THE ANNALS
of
MATHEMATICAL STATISTICS

(FOUNDED BY H. C. CARVER)

THE OFFICIAL JOURNAL OF THE INSTITUTE OF MATHEMATICAL STATISTICS

Contents

Vol. X, No. 3 — September, 1939

THE ANNALS OF MATHEMATICAL STATISTICS

Manuscripts for publication in the ANNALS OF MATHEMATICAL STATISTICS should be sent to S. S. Wilks, Fine Hall, Princeton, New Jersey. Manuscripts should be typewritten double-spaced with wide margins, and the original copy should be submitted. Footnotes should be reduced to a minimum and whenever possible replaced by a bibliography at the end of the paper; formulae in footnotes should be avoided. Figures, charts, and diagrams should be drawn on plain white paper or tracing cloth in black India ink twice the size they are to be printed. Authors are requested to keep in mind typographical difficulties of complicated mathematical formulae.

Authors will ordinarily receive only galley proofs. Fifty reprints without covers will be furnished free. Additional reprints and covers furnished at cost.

The subscription price for the ANNALS is $4.00 per year. Single copies $1.25. Back numbers are available at the following rates:

Vols. I-IV $5.00 each. Single numbers $1.50.
Vols. V to date $4.00 each. Single numbers $1.25.

Subscriptions, renewals, orders for back numbers and other business communications should be sent to A. T. Craig, University of Iowa, Iowa City, Iowa.

The ANNALS OF MATHEMATICAL STATISTICS is published quarterly by the Institute of Mathematical Statistics.

COMPOSED AND PRINTED AT THE
WAVERLY PRESS, INC.
BALTIMORE, MD., U. S. A.

ON THE SAMPLING THEORY OF ROOTS OF DETERMINANTAL EQUATIONS

By M. A. Girshick[1]

In a recent paper[2] Hotelling has considered two functions of the covariances of two sets of variates (having a multivariate normal distribution with s variates in the first set, t variates in the second, $s \leq t$) which he designates by Q and Z and which he defines as follows:

$$Q^2 = \frac{(-1)^s C}{AB} \quad \text{and}^3 \quad Z = \frac{D}{AB} \tag{1.1}$$

where A is the determinant of the covariances among the variates of the first set, B the determinant of the covariances among the variates of the second set, D the determinant of covariances of the two sets taken together, and C a determinant obtained from D by replacing the covariances among the variates of the first set by zeros. Both Q^2 and Z are shown to be invariant under internal linear transformations of either set of variates.

In solving the problem of determining linear functions of the two sets of variates for which the multiple correlation is a maximum, Hotelling arrives at a set of parameters $\rho_1, \rho_2, \cdots, \rho_s$ which he names "canonical correlations" and which are the positive or zero roots of the determinantal polynomial

$$D(\lambda) = \begin{vmatrix} -\lambda\sigma_{11} & \cdots & -\lambda\sigma_{1s} & \sigma_{1,s+1} & \cdots & \sigma_{1,s+t} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ -\lambda\sigma_{s1} & \cdots & -\lambda\sigma_{ss} & \sigma_{s,s+1} & \cdots & \sigma_{s,s+t} \\ \sigma_{s+1,1} & \cdots & \sigma_{s+1,s} & -\lambda\sigma_{s+1,s+1} & \cdots & -\lambda\sigma_{s+1,s+t} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \sigma_{s+t,1} & \cdots & \sigma_{s+t,s} & -\lambda\sigma_{s+t,s+1} & \cdots & -\lambda\sigma_{s+t,s+t} \end{vmatrix}. \tag{1.2}$$

The ρ's are equal in number to the variates of the first set and bear the following relations to Q and Z:

$$Q^2 = \rho_1^2\rho_2^2 \cdots \rho_s^2 \tag{1.3}$$

$$Z = (1 - \rho_1^2)(1 - \rho_2^2) \cdots (1 - \rho_s^2). \tag{1.4}$$

The corresponding functions for the sample covariances Hotelling designates by q and z, and the sample canonical correlations by $r_1, r_2, \cdots, r_s$. Under the assumption of complete independence between the two sets of variates and

---

[1] Most of this Research was accomplished at Columbia University under a Grant-in-Aid from the Carnegie Corporation of New York.

[2] Harold Hotelling, "Relations Between Two Sets of Variates," *Biometrika*, Vol. XXVIII, Dec. 1936.

[3] The function Z was first considered by S. S. Wilks in *Biometrika*, Vol. XXIV, Nov. 1932.

in the case $s = 2$ and $t = 2$, he shows that the joint distribution of q and z is of the form

$$\frac{1}{2}(n-2)(n-3)z^{\frac{1}{2}(n-5)}\,dq\,dz \tag{1.5}$$

q and z satisfying the inequalities

$$0 \leq z \leq 1, \qquad 0 \leq q \leq 1, \qquad z \leq (1-q)^2$$

and the joint distribution of the canonical correlations r_1 and r_2 is of the form

$$(n-2)(n-3)(r_1^2 - r_2^2)(1-r_1^2)^{\frac{1}{2}(n-5)}(1-r_2^2)^{\frac{1}{2}(n-5)}\,dr_1\,dr_2 \tag{1.6}$$

where n is one less than the number in the sample for each variate.

I

In Part I of this paper we shall, assuming independence between the two sets, find the joint moments of q and z for a general value of s and t and extend the joint distribution of q and z and hence of the canonical correlations to the case where there are two variates in the first set and any number of variates in the second, i.e. $s = 2$ and $t > 2$.[4]

1. **Joint Moments of q and z.** Since we are assuming complete independence between the two sets of variates we may without any loss of generality represent the sample values of the second set as points on the first t axes of unit distance from the origin in a space of n dimensions. The matrix of observations in the case of s variates in the first set and t variates in the second set will take the form

$$\left\| \begin{matrix} x_{11} & x_{12} & x_{13} \cdots & x_{1t} \cdots & x_{1n} \\ x_{21} & x_{22} & x_{23} \cdots & x_{2t} \cdots & x_{2n} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ x_{s1} & x_{s2} & x_{s3} \cdots & x_{st} \cdots & x_{sn} \\ 1 & 0 & 0 \cdots & 0 \cdots & 0 \\ 0 & 1 & 0 \cdots & 0 \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 \cdots & 1 \cdots & 0 \end{matrix} \right\| \tag{1.7}$$

The polynomial $D(\lambda)$ of (1.2) in terms of sample variances and covariances calculated from (1.7) then becomes

$$D(\lambda) = \begin{vmatrix} -\lambda a_{11} & \cdots & -\lambda a_{1s} & x_{11} & \cdots & x_{1t} \\ \cdot & & \cdot & \cdot & & \cdot \\ -\lambda a_{s1} & \cdots & -\lambda a_{ss} & x_{s1} & \cdots & x_{st} \\ x_{11} & \cdots & x_{s1} & -\lambda & \cdots & 0 \\ \cdot & & \cdot & \cdot & & \cdot \\ x_{1t} & \cdots & x_{st} & 0 & \cdots & -\lambda \end{vmatrix} \tag{1.8}$$

where $a_{ij} = \sum_1^n x_i x_j$.

[4] This extension is a generalization of Hotelling's method loc. cit.

We multiply the first s rows of (1.8) by λ and factor out λ from the last t columns. This yields

$$(1.9) \qquad D(\lambda) = \lambda^{t-s} \begin{vmatrix} -\lambda^2 a_{11} & \cdots & -\lambda^2 a_{1s} & x_{11} & \cdots & x_{1t} \\ \cdot & & \cdot & \cdot & & \cdot \\ -\lambda^2 a_{s1} & \cdots & -\lambda^2 a_{ss} & x_{s1} & \cdots & x_{st} \\ x_{11} & \cdots & x_{s1} & -1 & \cdots & 0 \\ \cdot & & \cdot & \cdot & & \cdot \\ x_{1t} & \cdots & x_{st} & 0 & \cdots & -1 \end{vmatrix}.$$

As a further simplification, we multiply the $(s + j)^{\text{th}}$ column by x_{ij} for all j from 1 to t and add the result to the i^{th} column. When this is done for every value of i from 1 to s and the resulting determinant expanded by means of the last t columns, the determinantal polynomial (1.9) becomes

$$D(\lambda) = \lambda^{t-s} \begin{vmatrix} b_{11} - \lambda^2 a_{11} & b_{12} - \lambda^2 a_{12} & \cdots & b_{1s} - \lambda^2 a_{1s} \\ \cdot & \cdot & \cdots & \cdot \\ b_{s1} - \lambda^2 a_{s1} & b_{s2} - \lambda^2 a_{s2} & \cdots & b_{ss} - \lambda^2 a_{ss} \end{vmatrix}$$

or symbolically

$$(1.10) \qquad D(\lambda) = \lambda^{t-s} \mid b_{ij} - \lambda^2 a_{ij} \mid$$

where $b_{ij} = \sum_1^t x_i x_j$.

Hence the s roots of $D(\lambda)$ which do not necessarily vanish may be obtained from the polynomial

$$(1.11) \qquad Q(\lambda) = \mid b_{ij} - \lambda^2 a_{ij} \mid.$$

The coefficient of the highest power of λ in $Q(\lambda)$ is given by $\mid a_{ij} \mid$, the determinant of the elements a_{ij}. Taking this in conjunction with (1.3) and (1.4) we see that

$$(1.12) \qquad \begin{aligned} q^2 &= \frac{Q(0)}{\mid a_{ij} \mid} = \frac{\mid b_{ij} \mid}{\mid a_{ij} \mid} \\ z &= \frac{Q(1)}{\mid a_{ij} \mid} = \frac{\mid c_{ij} \mid}{\mid a_{ij} \mid} \end{aligned}$$

where $c_{ij} = \sum_{t+1}^n x_i x_j$.

From the equations (1.12) we obtain

$$(1.13) \qquad E\{\mid a_{ij} \mid^{\frac{1}{2}(\alpha+2\beta)} q^\alpha z^\beta\} = E\{\mid b_{ij} \mid^{\frac{1}{2}\alpha} \mid c_{ij} \mid^\beta\}$$

where E stands for the mathematical expectation of the expressions in the $\{ \quad \}$.

It is obvious from the definition of b_{ij} and c_{ij} that the two determinants $\mid b_{ij} \mid$ and $\mid c_{ij} \mid$ are independently distributed. Moreover, the joint distribution of q and z does not depend on the determinant $\mid a_{ij} \mid$. The truth of the latter statement can be seen from the following geometrical considerations. If we con-

sider the sample values of each variate as a point in an n-dimensional space, then the two sets of variates determine two flat spaces, one of s dimensions and one of t dimensions in that space. A sample canonical correlation can then be considered as the cosine of a certain minimum or stationary angle between two lines, one line lying in the flat s space and the other in the flat t space. Since q and z are functions of the canonical correlations, they therefore depend only on lines and angles between two planes. The quantities a_{ij} on the other hand, depend on lines and angles lying entirely within one of these planes.

From the above considerations we see that equation (1.13) can be written as

$$E\{|a_{ij}|^{\frac{1}{2}(\alpha+2\beta)}\}E(q^{\alpha}z^{\beta}) = E(|b_{ij}|^{\frac{1}{2}\alpha})E(|c_{ij}|^{\beta})$$

or

$$E(q^{\alpha}z^{\beta}) = \frac{E(|b_{ij}|^{\frac{1}{2}\alpha})E(|c_{ij}|^{\beta})}{E(|a_{ij}|^{\frac{1}{2}(\alpha+2\beta)})} \tag{1.14}$$

The m^{th} moment of a determinant $|d_{ij}|$ of sums of sample cross products of p variates is given by the formula[5]

$$E(|d_{ij}|^{m}) = \frac{2^{pm}}{|D_{ij}|^{m}} \prod_{i=1}^{p} \left[\frac{\Gamma\left(\frac{n+2m+1-i}{2}\right)}{\Gamma\left(\frac{n+1-i}{2}\right)} \right], \tag{1.15}$$

where D_{ij} denotes the cofactor corresponding to σ_{ij} divided by the determinant $|\sigma_{ij}|$. Substituting (1.15) in (1.14) and simplifying, we get for the joint moments of q and z

$$E(q^{\alpha}z^{\beta}) = \prod_{i=1}^{s} \left[\frac{\Gamma\left(\frac{t+\alpha+1-i}{2}\right)\Gamma\left(\frac{n-t+2\beta+1-i}{2}\right)\Gamma\left(\frac{n+1-i}{2}\right)}{\Gamma\left(\frac{t+1-i}{2}\right)\Gamma\left(\frac{n-t+1-i}{2}\right)\Gamma\left(\frac{n+\alpha+2\beta+1-i}{2}\right)} \right]. \tag{1.16}$$

2. Joint Distribution of q and z for $s = 2$, $t > 2$. In order to determine the joint distribution of q and z for $s = 2$ and $t > 2$, we shall first prove the following lemma.

LEMMA: *Let q and z be defined as in* (1.1) *for two sets of variates having s variates in either set and let q' and z' be similarly defined with $s < t$ where s is the number of variates in the first set and t the number of variates in the second set, then for $n = t + s$, the joint distribution of q^2 and z is identical with that of z' and q'^2.*

PROOF. If the number of variates in either set are the same and $n = t + s$, then by (1.12)

$$q^2 = \frac{|b_{ij}|}{|a_{ij}|}, \qquad z = \frac{|c_{ij}|}{|a_{ij}|}$$

[5] Cf. S. S. Wilks, "Certain Generalizations in the Analysis of Variance," *Biometrika*, Vol. XXIV, Nov. 1932.

where

$$(1.17) \qquad b_{ij} = \sum_{1}^{t} x_i x_j, \qquad c_{ij} = \sum_{s+1}^{t+s} x_i x_j, \qquad a_{ij} = \sum_{1}^{t+s} x_i x_j$$

and $s = t$.

However, for $s < t$, and $n = t + s$, we take for the second set of t variates points on the t axes at unit distance from the origin in the $(t + s)$-dimensional space *perpendicular* to the first s axes. The matrix of observations in this case takes the form

$$(1.18) \qquad \left\| \begin{matrix} x_{11} & x_{12} & \cdots & x_{1s} & x_{1,s+1} & \cdots & x_{1,s+t} \\ x_{21} & x_{22} & \cdots & x_{2s} & x_{2,s+1} & \cdots & x_{2,s+t} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdots & \cdot \\ x_{s1} & x_{s2} & \cdots & x_{ss} & x_{s,s+1} & \cdots & x_{s,s+t} \\ 0 & 0 & \cdots & 0 & 1 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 1 \end{matrix} \right\|.$$

Employing the same arguments as in equations (1.8) (1.9) and (1.10) we find that

$$(1.19) \qquad Q(\lambda) = | c_{ij} - \lambda^2 a_{ij} |, \qquad q'^2 = \frac{| c_{ij} |}{| a_{ij} |}, \qquad z' = \frac{| b_{ij} |}{| a_{ij} |}$$

where

$$b_{ij} = \sum_{1}^{t} x_i x_j, \qquad c_{ij} = \sum_{s+1}^{t+s} x_i x_j, \qquad a_{ij} = \sum_{1}^{t+s} x_i x_j.$$

Comparing these equations with (1.17) we see that

$$(1.20) \qquad z = q'^2, \qquad q^2 = z'.$$

This proves the lemma.

Now let $s = 2$. Setting $n = t + 2$ in equation (1.5) and using the transformation (1.20) we get for the joint distribution of q' and z'

$$(1.21) \qquad \tfrac{1}{2}t(t-1)q'^{t-2}z'^{-\frac{1}{2}}\, dq'\, dz'.$$

Let r be the correlation between the two variates of the first set. The distribution of r in samples for which $n = t + 2$ when the population correlation is zero is known to be

$$(1.22) \qquad \frac{\Gamma\left(\frac{t+2}{2}\right)}{\Gamma\left(\frac{t+1}{2}\right)\sqrt{\pi}} (1 - r^2)^{\frac{1}{2}(t-1)}\, dr.$$

The distribution of r is independent of q and z. Hence, the joint distribution of q', z', and r is given by the product of (1.21) and (1.22). Dropping the

primes from q' and z' in (1.21), we get for the joint distribution of the three quantities in the case $n = t + 2$,

$$\frac{1}{2} t(t-1) \frac{\Gamma\left(\frac{t+2}{2}\right)}{\Gamma\left(\frac{t+1}{2}\right)\sqrt{\pi}} q^{t-2} z^{-\frac{1}{2}} (1 - r^2)^{\frac{1}{2}(t-1)} \, dq \, dz \, dr. \tag{1.23}$$

We shall now derive the joint distribution of q and z for a general value of n for $s = 2$, $t > 2$. We set $x_1 = x$, $x_2 = y$ and take the t sample variates of the second set to be points on the first t axes at unit distance from the origin in a space of n dimensions. As in (1.12) calculate q and z.

$$q^2 = \frac{\sum_1^t x^2 \sum_1^t y^2 - \left(\sum_1^t xy\right)^2}{1 - r^2}, \quad z = \frac{\sum_{t+1}^n x^2 \sum_{t+1}^n y^2 - \left(\sum_{t+1}^n xy\right)^2}{1 - r^2}. \tag{1.24}$$

We transform the points $(x_1, \cdots, x_n)$ and $(y_1, \cdots, y_n)$ to hyperspherical coordinates, the transformation to be represented parametrically by the equations

$$\begin{aligned}
x_1 &= \sin\theta_1 \sin\theta_2 \cdots \sin\theta_{t-1} \sin\theta_t \\
x_2 &= \cos\theta_1 \sin\theta_2 \cdots \sin\theta_{t-1} \sin\theta_t \\
x_3 &= \cos\theta_2 \cdots \sin\theta_{t-1} \sin\theta_t \\
&\cdots \\
x_t &= \cos\theta_{t-1} \sin\theta_t \\
x_{t+1} &= \cos\theta_t \cos\theta_{t+1} \\
x_{t+2} &= \cos\theta_t \sin\theta_{t+1} \cos\theta_{t+2} \\
&\cdots \\
x_{n-1} &= \cos\theta_t \sin\theta_{t+1} \sin\theta_{t+2} \cdots \cos\theta_{n-1} \\
x_n &= \cos\theta_t \sin\theta_{t+1} \sin\theta_{t+2} \cdots \sin\theta_{n-1}
\end{aligned} \tag{1.25}$$

with the same representation for the y's in terms of parameters $\phi_1, \phi_2, \cdots \phi_{n-1}$. It is to be observed that in (1.24) and (1.25) $\Sigma x^2 = 1$, $\Sigma y^2 = 1$. This we may assume since q and z are invariant under such transformations.

In this new coordinate system, our samples $(x_1, \cdots, x_n)$ and $(y_1, \cdots, y_n)$ are taken as random points on a unit hypersphere about the origin in n dimensions. There is no loss of generality in this since x and y are assumed to be uncorrelated in the population and hence possess spherical symmetry of the density distribution in a space of n dimensions.

The element of probability for the x points on this hypersphere is proportional to the $(n-1)$-dimensional area on this sphere. Now the $n - 1$ dimensional area is given by

$$\sqrt{g} \, d\theta_1 \, d\theta_2 \cdots d\theta_{n-1}$$

where g is a determinant of order $n - 1$ in which the element in the i^{th} row and j^{th} column is

$$\sum_{\alpha=1}^{n} \frac{\partial x_\alpha \partial x_\alpha}{\partial \theta_i \partial \theta_j}.$$

When $i \neq j$, all these quantities vanish as can be seen by inspection from (1.25). When $i = j$, we have

$$\sum_1^n \left(\frac{\partial x_\alpha}{\partial \theta_1}\right)^2 = \sin^2 \theta_2 \sin^2 \theta_3 \cdots \sin^2 \theta_t$$

$$\sum_1^n \left(\frac{\partial x_2}{\partial \theta_2}\right)^2 = \sin^2 \theta_3 \cdots \sin^2 \theta_t$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\sum_1^n \left(\frac{\partial x_\alpha}{\partial \theta_t}\right)^2 = 1$$

$$\sum_1^n \left(\frac{\partial x_\alpha}{\partial \theta_{t+1}}\right)^2 = \cos^2 \theta_t$$

$$\sum_1^n \left(\frac{\partial x_\alpha}{\partial \theta_{t+2}}\right)^2 = \cos^2 \theta_t \sin^2 \theta_{t+1}$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\sum_1^n \left(\frac{\partial x_\alpha}{\partial \theta_{n-1}}\right)^2 = \cos^2 \theta_j \sin^2 \theta_{j+1} \cdots \sin^2 \theta_{n-2}.$$

Therefore

$$g = \sin^2 \theta_2 \sin^4 \theta_3 \cdots \sin^{2(t-1)} \theta_t \cos^{2(n-t-1)} \theta_t \cdots \sin^2 \theta_{n-2}$$

and hence the element of generalized area is given by

$$\begin{aligned}(1.26)\qquad &\sin \theta_2 \sin^2 \theta_3 \cdots \sin^{t-1} \theta_t \cos^{n-t-1} \theta_t \\ &\qquad \sin^{n-t-2} \theta_{t+1} \cdots \sin \theta_{n-2}\, d\theta_1\, d\theta_2 \cdots d\theta_{n-1}.\end{aligned}$$

Similarly we can show that the element of generalized area for the y point is

$$\begin{aligned}(1.27)\qquad &\sin \phi_2 \sin^2 \phi_3 \cdots \sin^{t-1} \phi_t \cos^{n-t-1} \phi_t \\ &\qquad \sin^{n-t-2} \phi_{t+1} \cdots \sin \phi_{n-2}\, d\phi_1\, d\phi_2 \cdots d\phi_{n-1}.\end{aligned}$$

The joint distribution of $\theta_1, \theta_2, \cdots, \theta_{n-1}$ and $\phi_1, \phi_2, \cdots, \phi_{n-1}$ (since the θ's are independent of the ϕ's) is proportional to the product of (1.26) and (1.27).

We now introduce four new sets of variables, u, v, u', v', defined by the following equations

$$(1.28)\qquad x_i = u_i \sin \theta_t, \qquad y_i = v_i \sin \phi_t \qquad (i = 1, 2, \cdots, t)$$

$$(1.29)\qquad x_j = u_j' \cos \theta_t \qquad y_j = v_j' \cos \phi_t \qquad (j = t + 1, \cdots, n).$$

The u_i and v_i can be regarded as two points on a sphere in a space of t dimensions and u_j' and v_j' as two points on a sphere in a space of $n - t$ dimensions.

Let λ be the angle between the two points u and v and μ the angle between the two points u' and v'. Then

$$\cos \lambda = \sum_{i=1}^{t} u_i v_i ; \qquad \cos \mu = \sum_{j=t+1}^{n} u'_j v'_j .$$

The probability element for λ is proportional to $\sin^{t-2} \lambda \, d\lambda$, and that for μ is proportional to $\sin^{n-t-2} \mu \, d\mu$.

From the definition of u_i and v_i, we see that they depend only on $\theta_1, \theta_2, \cdots, \theta_{t-1}$; $\phi_1, \phi_2, \cdots, \phi_{t-1}$ respectively, and u'_j and v'_j depend only on $\theta_{t+1}, \theta_{t+2}, \cdots, \theta_{n-1}$; $\phi_{t+1}, \phi_{t+2}, \cdots, \phi_{n-1}$ respectively. It follows that the quantities λ, μ, θ_t, and ϕ_t are independently distributed.

The joint distribution of the θ's and ϕ's we integrate between constant limits with respect to all the variates except θ_t and ϕ_t. This gives for the joint distribution of θ_t and ϕ_t

$$A_n \sin^{t-1} \theta_t \sin^{t-1} \phi_t \cos^{n-t-1} \theta_t \cos^{n-t-1} \phi_t \, d\theta_t \, d\phi_t$$

where A_n is a constant depending only on n.

Multiplying this by the distributions of λ and μ and dropping the subscript t from θ and ϕ we get for the joint distribution of λ μ, θ, and ϕ

$$(1.30) \qquad k_n \sin^{t-1} \theta \sin^{t-1} \phi \cos^{n-t-1} \theta \cos^{n-t-1} \phi \sin^{t-2} \lambda \cos^{n-t-2} \mu \, d\theta \, d\phi \, d\lambda \, d\mu$$

where k_n is a constant depending on n. The limits of integration for θ and ϕ are 0 and $\pi/2$; for λ and μ they are 0 and π.

Expressing q and z in terms of the new quantities as defined in (1.25), (1.28) and (1.29) we get

$$(1.31) \qquad q^2 = \frac{\left(\sum_1^t x^2\right)\left(\sum_1^t y^2\right) - \left(\sum_1^t xy\right)^2}{1 - r^2} = \frac{\sin^2 \theta \sin^2 \phi \sin^2 \lambda}{1 - r^2}$$

$$(1.32) \qquad z = \frac{\left(\sum_{t+1}^n x^2\right)\left(\sum_{t+1}^n y^2\right) - \left(\sum_{t+1}^n xy\right)^2}{1 - r^2} = \frac{\cos^2 \theta \cos^2 \phi \sin^2 \mu}{1 - r^2}$$

where

$$(1.33) \qquad r = \Sigma xy = \sin \theta \sin \phi \cos \lambda + \cos \theta \cos \phi \cos \mu$$

is the sample correlation between x and y.

We now consider a transformation of the variables θ, ϕ, and μ in (1.30) to the new variables q, z, and r. Without troubling to compute the Jacobian J of the transformation, we know that it is independent of n since the relations (1.31), (1.32) and (1.33) do not involve n. Substituting from (1.31) and (1.32) into (1.30) we get for the joint distribution of q, z, r, and λ

$$k_n \psi q^{t-1} z^{\frac{1}{2}(n-t-1)} (1 - r^2)^{\frac{1}{2}(n-2)} \, dq \, dz \, dr \, d\lambda$$

where ψ is independent of n. Integrating with respect to λ between limits which are independent of n, we get for the joint distribution of q, z, and r

$$k_n \Psi q^{t-1} z^{\frac{1}{2}(n-t-1)} (1 - r^2)^{\frac{1}{2}(n-2)}\, dq\, dz\, dr. \tag{1.34}$$

But, for $n = t + 2$, this joint distribution reduces to (1.23). Therefore

$$k_{t+2}\Psi = \frac{1}{2}\, t(t-1) \frac{\Gamma\left(\frac{t+2}{2}\right)}{\Gamma\left(\frac{t+1}{2}\right)\sqrt{\pi}}\, z^{-1} q^{-1} (1 - r^2)^{-\frac{1}{2}}$$

so that (1.34) can be written as

$$k'_n q^{t-2} z^{\frac{1}{2}(n-t-3)} (1 - r^2)^{\frac{1}{2}(n-3)}\, dq\, dz\, dr.$$

However, since the distribution of r is known to be

$$\frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)\sqrt{\pi}} (1 - r^2)^{\frac{1}{2}(n-3)}\, dr$$

we finally get for the joint distribution of q and z

$$h_n q^{t-2} z^{\frac{1}{2}(n-t-3)}\, dq\, dz$$

where h_n depends on n. The integral over the entire region defined by the inequalities

$$0 \leq q \leq 1, \qquad 0 \leq z \leq 1, \qquad z \leq (1 - q)^2$$

must equal unity; the constant h_n is therefore readily found to be $\dfrac{(n-2)!}{2(t-2)!\,(n-t-2)!}$. Thus the joint distribution in the final form is

$$\frac{(n-2)!}{2(t-2)!\,(n-t-2)!}\, q^{t-2} z^{\frac{1}{2}(n-t-3)}\, dq\, dz\,. \tag{1.35}$$

Now by (1.3) and (1.4), $q = r_1 r_2$, $z = (1 - r_1^2)(1 - r_2^2)$, and hence the Jacobian

$$\frac{\partial(q, z)}{\partial(r_1, r_2)} = 2(r_1^2 - r_2^2)\,. \tag{1.36}$$

Making the transformation in (1.35) we get the joint distribution of the canonical correlations r_1 and r_2 (for the case $s = 2$ and a general value of t) in the form

$$\frac{(n-2)!}{(t-2)!\,(n-t-2)!} (r_1^2 - r_2^2)(r_1 r_2)^{t-2} \left[(1 - r_1^2)(1 - r_2^2)\right]^{\frac{1}{2}(n-t-3)}\, dr_1\, dr_2\,. \tag{1.37}$$

II. JOINT LIMITING DISTRIBUTIONS OF CANONICAL CORRELATIONS AND LATENT ROOTS

In formula (1.37) we set

$$k_1 = nr_1^2, \qquad k_2 = nr_2^2$$

and get for the joint distribution of k_1 and k_2

$$\text{(2.1)} \qquad \frac{(n-2)!}{4(t-2)!\,(n-t-2)!\,n^t}(k_1 - k_2)(k_1k_2)^{\frac{1}{2}(t-3)} \left[\left(1 - \frac{k_1}{n}\right)\left(1 - \frac{k_2}{n}\right)\right]^{\frac{1}{2}(n-t-3)} dk_1\,dk_2.$$

When $n \to \infty$, the quantity $\frac{(n-2)!}{n^t(n-t-2)!}$ approaches 1 and $\left(1 - \frac{k}{n}\right)^{\frac{1}{2}(n-t-3)}$ approaches $e^{-\frac{1}{2}k^2}$. Hence the limiting distribution of the two canonical correlations is given by

$$\text{(2.2)} \qquad \frac{1}{4(t-2)!}(k_1 - k_2)(k_1k_2)^{\frac{1}{2}(t-3)}\, e^{-\frac{1}{2}(k_1+k_2)}\, dk_1\,dk_2.$$

We shall call (2.2) the "generalized chi-square" distribution and show that the roots of the characteristic polynomial

$$\text{(2.3)} \qquad \varphi(k) = \begin{vmatrix} a_{11} - k & a_{12} \\ a_{21} & a_{22} - k \end{vmatrix}$$

are distributed in precisely this form. Here $a_{ij} = \Sigma x_i x_j$ where x_1 and x_2 are normally and independently[6] distributed with unit variance in the population and zero mean in the sample.

Let k_1 and k_2 be the roots of (2.3). That is, k_1 and k_2 are the two roots of the quadratic equation

$$\text{(2.4)} \qquad k^2 - p_1k + p_2 = 0$$

where

$$\text{(2.5)} \qquad p_1 = k_1 + k_2 = a_{11} + a_{22}$$

$$\text{(2.6)} \qquad p_2 = k_1k_2 = a_{11}a_{22} - a_{12}^2.$$

In the absence of correlation in the population, the joint distribution of a_{11}, a_{22} and a_{12} is known to be

$$\text{(2.7)} \qquad h_n \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}^{\frac{1}{2}(n-3)} e^{-\frac{1}{2}(a_{11}+a_{22})}\, da_{11}\,da_{22}\,da_{12}$$

where h_n is a constant depending only on n.

[6] The part of the assumption relating to independence may be removed without loss of generality. See last paragraph below.

We consider a transformation to the variables p_1, p_2 and a_{12}. From (2.5) and (2.6) we calculate the Jacobian J of the transformation,

$$J = \frac{1}{a_{11} - a_{22}} \tag{2.8}$$

and since

$$2a_{ii} = p_1 \pm (p_1^2 - 4p_2 - 4a_{12}^2)^{\frac{1}{2}} \tag{2.9}$$

$$J = \frac{1}{(p_1^2 - 4p_2 - 4a_{12}^2)^{\frac{1}{2}}}. \tag{2.10}$$

Substituting from (2.5) and (2.6) into (2.7) and multiplying by J, we get for the joint distribution of k_1, k_2 and a_{12}

$$h_n p_2^{\frac{1}{2}(n-3)} e^{-\frac{1}{2}p_1} \frac{dp_1\, dp_2\, da_{12}}{(p_1^2 - 4p_2 - 4a_{12}^2)^{\frac{1}{2}}}. \tag{2.11}$$

We make the transformation $u = a_{12}^2$ and get for the joint distribution of k_1, k_2 and u

$$\frac{h_n}{2} p_2^{\frac{1}{2}(n-3)} e^{-\frac{1}{2}p_1} \frac{dp_1\, dp_2\, du}{(bu - 4u^2)^{\frac{1}{2}}} \tag{2.12}$$

where $b = p_1^2 - 4p_2$.

Since both a_{11} and a_{22} are real, equation (2.9) shows that $b - 4u \geq 0$. Hence the limits of integration for u are 0 and $\frac{b}{4}$. Integrating out u in (2.12) between the above limits we obtain the joint distribution of p_1 and p_2.

Now the integral

$$\int_0^{b/4} \frac{du}{(bu - 4u^2)^{\frac{1}{2}}} = -\frac{1}{2}\sin^{-1}\left(\frac{-8u + b}{b}\right)\Bigg]_0^{b/4} = c \tag{2.13}$$

where c is some constant. Hence the joint distribution of p_1 and p_2 is given by

$$H_n p_2^{\frac{1}{2}(n-3)} e^{-\frac{1}{2}p_1}\, dp_1\, dp_2. \tag{2.14}$$

By integrating (2.14) over the region $0 \leq p_2 \leq \left(\frac{p_1}{2}\right)^2$ and $0 \leq p_1 \leq \infty$ we get $H_n = \frac{1}{4}(n - 2)!$.

We next transform p_1 and p_2 in terms of k_1 and k_2 from (2.5) and get for the joint distribution of k_1 and k_2

$$\frac{1}{4(n-2)!}(k_1 - k_2)(k_1 k_2)^{\frac{1}{2}(n-3)} e^{-\frac{1}{2}(k_1+k_2)}\, dk_1\, dk_2. \tag{2.15}$$

This distribution is identical with that of (2.2) with $n = t$.

The above is an example of a more general

THEOREM: *Let $r_1, r_2, \cdots, r_s$ be a set of simple finite canonical roots of the two independent sets of variates $x_1, \cdots, x_s$, and $x_{s+1}, \cdots, x_{s+t}$. Let $k_i = nr_i^2$ $(i = 1, 2, \cdots, s)$. Then the joint limiting distribution of the k's approaches the exact joint sampling distribution of the latent roots of a matrix of sample product sums with t degrees of freedom of s normally distributed variates having unit variance in the population.*

PROOF: The proof follows from equation (1.11). For let us multiply and divide a_{ij} in (1.11) by n and set $n\lambda^2 = k$. The determinantal polynomial becomes

$$\varphi(k) = | b_{ij} - ks_{ij} |. \tag{2.16}$$

Without any loss of generality, we so transform the first set of variates that they become of zero correlation and unit variance in the population. Then it follows that

$$E(s_{ij}) = E\left(\sum \frac{x_i x_j}{n}\right) = \delta_{ij}$$

where δ_{ij} equals zero for $i \neq j$ and 1 for $i = j$.

Now let $P(x \geq a)$ stand for the probability that the variate x be greater than or equal to some constant a. Then, by the Strong Law of Large Numbers we can state that, given an $\epsilon > 0$ and a $\delta > 0$ there exists a positive integer n_0 such that for $n > n_0$

$$P\{| s_{ij} - \delta_{ij}| \geq \delta\} \leq \epsilon.$$

If then we let n increase indefinitely, the quantity $b_{ij} = \sum_1^t x_i x_j$ remains fixed while s_{ij} approaches, in the probability sense, δ_{ij}. Since the roots of a polynomial are continuous functions of the coefficients, we can, by an extension of the Law of Large Numbers, show that in the limit the roots of (2.16) will be distributed like the roots of the polynomial

$$\varphi(k) = | b_{ij} - k\delta_{ij} |.$$

This proves the theorem.

COROLLARY 1. *The limiting distribution of q^2 in case of complete independence between the two sets of variates approaches the exact distribution of a generalized sample variance (i.e. a determinant of sample variances and covariances) with t degrees of freedom. The proof follows from the fact that q^2 is a product of the roots of* (1.11) *and therefore by the above theorem, is distributed in the limit like* $| b_{ij} |$.

COROLLARY 2. *The distribution of the sum of the squares of the canonical correlations approach in the limit a χ^2 distribution with st degrees of freedom. This is obvious since in the limit the sum of the squares of the roots, by the above theorem, has the distribution of $b_{11} + b_{22} + \cdots + b_{ss}$ and each b_{jj} is distributed like χ^2 with t degrees of freedom.*

While the canonical roots of (1.2) are invariant under any non-singular linear transformations, the latent roots of a determinant of sample covariances are

invariant only under an orthogonal transformation. But there exists an orthogonal transformation which reduces a set of variates having a multivariate normal distribution to a set which are normally and independently distributed with variances equal to the latent roots of the population generalized variances of the original variates. Hence, in dealing with the distribution of latent roots, we may assume independence in the population without any loss of generality but the assumption of equal variance leads only to a special case. Moreover, the above consideration also explains the form of the asymptotic error of the sample latent root given in Part III of this paper.

III. ASYMPTOTIC STANDARD ERRORS OF LATENT ROOTS AND COEFFICIENTS OF PRINCIPAL COMPONENTS

1. Many statisticians have had occasion to use in their statistical analyses characteristic roots (or as they are sometimes called "latent" roots) of determinants of correlations or covariances. Especially has this become true since the publication of Hotelling's paper on principal components.[7] It is therefore of great importance to find, if not their sampling distributions, at least their limiting distributions and their asymptotic standard errors. This we shall do in this paper for the case of non-vanishing simple roots and by the same method[8] get the asymptotic variances and covariances of the coefficients of principal components. We have already derived in Part II the sampling distribution of the two latent roots of a determinant of covariances obtained from two normally distributed variates having equal variance in the population. This distribution is of no great importance in itself except that it gives us some idea as to the form of the distribution in the general case.

In what follows, we shall use the convention that a repeated subscript in the same term stands for summation. If repeated subscripts appearing in a term are not to be summed, we shall place them in brackets following the expression in which they appear. Thus in the equation (3.1) below, we sum with respect to j but not with respect to q even though on the right hand side q appears twice.

Let $x_1, x_2, \cdots, x_s$ be a set of variates which have a multi-variate normal distribution. We assume that these variates have been resolved into components by Hotelling's method.[9] Let $\gamma_1, \gamma_2, \cdots, \gamma_s$ be the principal components. Then $x_i = a_{ij}\gamma_j$. The a_{ij}'s satisfy the following equations:

$$a_{jq}\sigma_{ij} = \lambda_q a_{iq}, \; [q] \tag{3.1}$$

$$a_{ip}a_{iq} = \lambda_q \delta_{pq} \tag{3.2}$$

---

[7] "Analysis of a Complex of Statistical Variables into Principal Components," *The Journal of Educational Psychology*, Sept. and Oct. 1933. See also M. A. Girshick, "Principal Components," *Journal of the American Statistical Association*, Vol. 31, Sept. 1936.

[8] The method here employed is parallel to the one used by Hotelling in his paper of 1936 in deriving asymptotic standard errors for canonical correlations.

[9] Loc. cit.

where the symbol δ_{pq} has the value zero for $p \neq q$ and 1 for $p = q$, λ_q is a root of the characteristic equation

$$(3.3) \qquad | \sigma_{ij} - \lambda\delta_{ij} | = 0$$

and σ_{ij} is the population covariance of x_i and x_j.

If we multiply (3.1) by a_{ip}, sum with respect to i and use (3.2), we get

$$(3.4) \qquad a_{ip}a_{jq}\sigma_{ij} = \lambda_p^2\delta_{pq}.$$

When a root of (3.3) is simple and not equal to zero, the corresponding a_{ij}'s and the root itself are definite analytic functions of the σ_{ij}'s over a region without singularities. A set of sampling errors $d\sigma_{ij}$ in the covariances will then determine a corresponding set of sampling errors in the a_{ij}'s and in the root.

We assume then, that the roots $\lambda_1, \lambda_2, \cdots, \lambda_t$ of (3.3) we are considering are simple and non-vanishing. In terms of the derivatives of the analytic functions we define

$$(3.5) \qquad da_{rk} = \frac{\partial a_{rk}}{\partial \sigma_{pq}} d\sigma_{pq}, \qquad d\lambda_r = \frac{\partial \lambda_r}{\partial \sigma_{pq}} d\sigma_{pq}$$

where $d\sigma_{pq} = s_{pq} - \sigma_{pq}$, s_{pq} being the corresponding sample covariance.

Differentiating equation (3.1) and employing the above formulae we get

$$(3.6) \qquad \sigma_{ij}da_{jq} + a_{jq}d\sigma_{ij} = \lambda_q da_{iq} + a_{iq}d\lambda_q. \qquad [q]$$

We now multiply this equation by a_{ip}, sum with respect to i, and use equations (3.1) and (3.2). This yields:

$$(3.7) \qquad \lambda_p a_{jp}da_{jq} + a_{ip}a_{jq}d\sigma_{ij} = \lambda_q a_{ip}da_{iq} + \lambda_q\delta_{pq}d\lambda_q. \qquad [p, q]$$

When $p = q$, the term $\lambda_p a_{jp}da_{jp}$ cancels out and equation (3.7) reduces to

$$(3.8) \qquad \lambda_p d\lambda_p = a_{ip}a_{jp}d\sigma_{ij}. \qquad [p]$$

We change the subscripts p, i, j, to q, k, m, in (3.8) and multiply together the two equations thus obtained. This gives:

$$(3.9) \qquad \lambda_p\lambda_q d\lambda_p d\lambda_q = a_{ip}a_{jp}a_{kq}a_{mq}d\sigma_{ij}d\sigma_{km}. \qquad [p, q]$$

Hence

$$(3.10) \qquad \lambda_p\lambda_q E(d\lambda_p d\lambda_q) = a_{ip}a_{jp}a_{kq}a_{mq}E(d\sigma_{ij}d\sigma_{km}) \qquad [p, q]$$

where the symbol E denotes the mathematical expectation or mean value of the expression following.

Now it can be easily shown by means of the characteristic function of a multivariate normal distribution that

$$(3.11) \qquad E(d\sigma_{ij}d\sigma_{km}) = \frac{1}{n}(\sigma_{ik}\sigma_{jm} + \sigma_{im}\sigma_{jk})$$

where n is one less than the number in the sample. Substituting this expression in equation (3.10) and using (3.4) we get the following rather simple result

$$\lambda_p \lambda_q E(d\lambda_p d\lambda_q) = \frac{2\lambda_p^4 \delta_{pq}}{n}. \qquad [p, q] \tag{3.12}$$

Setting $p = q$ in this formula we get

$$E[(d\lambda_p)^2] = \frac{2\lambda_p^2}{n}. \tag{3.13}$$

But when $p \neq q$

$$E[d\lambda_p d\lambda_q] = 0. \tag{3.14}$$

Let $l_1, l_2, \cdots, l_t$, be the corresponding latent roots of a determinant of sample covariances. The sample latent root l_p may be expanded about λ_p in a Taylor series of the form

$$l_p = \lambda_p + \frac{\partial \lambda_p}{\partial \sigma_{rl}} d\sigma_{rl} + \frac{1}{2} \frac{\partial^2 \lambda_p}{\partial \sigma_{rl} \partial \sigma_{uv}} d\sigma_{rl} d\sigma_{uv} + \cdots \tag{3.15}$$

or, by (3.5)

$$l_p - \lambda_p = d\lambda_p + \cdots. \tag{3.16}$$

Squaring both sides of (3.16), taking the expected value, and using (3.13) we find *that the sample variance of a latent root* l_p, *apart from terms of higher order in* n^{-1}, *is given by* $\frac{2\lambda_p^2}{n}$.

If in (3.11) we set $i = j = k = m$, we get the variance of a sample variance, and it is interesting to note that its form is identical with the first term of the asymptotic expansion of the variance of a sample latent root.

The sample covariance of any two distinct roots is by (3.14) zero for the first term of the asymptotic expansion. That is, the covariance is at least of order n^{-2}. All the above results also follow from the fact, shown by the author in a previous paper,[10] that the coefficients of the principal components and hence the latent roots are maximum likelihood statistics. This property of the latent roots permits us also to state the following

THEOREM: *Let* $\lambda_1, \lambda_2, \cdots, \lambda_t$ *be any set of simple non-vanishing roots of* (3.3). *For sufficiently large samples these will be approximated by certain of the latent roots* $l_1, l_2, \cdots, l_t$ *of the samples. If* $l_i - \lambda_i$ *is divided by the standard error*

$$\sigma_{l_i} = \lambda_i \sqrt{\frac{2}{n}}$$

the resulting variates have a distribution which, as n increases, approaches the normal distribution of t independent variates of zero mean and unit standard deviation.

---

[10] Loc. cit.

COROLLARY: Let λ_1 be a maximum simple, non-vanishing root of (3.3) and let l_1 be the corresponding maximum sample root. Then, $l_1 - \lambda_1$ divided by its standard error has a distribution approaching normality in the limit.

2. **The Variance of Log l.** The formula for the standard error of the latent root given above contains a population parameter λ the numerical value of which we usually do not know. It is therefore important to find a transformation of the latent root to a new variate which will have or its leading term of the asymptotic standard error a quantity independent of the population parameter. Let $k = f(l)$ be such a transformation. Then $K = f(\lambda)$ is the corresponding transformation for the population root.

We now expand k in a Taylor series about $l = \lambda$

$$dk = f'(\lambda)dl + \tfrac{1}{2}f''(\lambda)(dl)^2 + \cdots \tag{3.17}$$

and get an approximation

$$dk = f'(\lambda)dl. \tag{3.18}$$

Squaring both sides and taking the expectation, we get

$$E(dk)^2 = [f'(\lambda)]^2 E[(dl)^2] = [f'(\lambda)]^2 \frac{2\lambda^2}{n}. \tag{3.19}$$

Now set $E(dk)^2 = 2/n$. Then, from (3.19)

$$f'(\lambda) = 1/\lambda$$

or

$$f(\lambda) = \log \lambda \tag{3.20}$$

Hence, if we set $k = \log l$, then

$$\sigma_k^2 = 2/n \tag{3.21}$$

is an approximation to the variance of k and is independent of any population parameter.

3. **The Asymptotic Variances and Covariances of Roots of Determinants of Correlations.** While the formulas for the asymptotic standard errors of the latent roots of a determinant of covariances are rather simple, this is not the case with the roots of a determinant of correlations. In deriving the asymptotic standard errors of simple non-vanishing roots of a determinant of correlations, we again assume that the variates $x_1, x_2, \cdots, x_s$, which in this case are of unit variance in the population, have been resolved into principal components. The equations of the previous section, up to and including (3.10), remain the same except that we substitute ρ_{ij} for every σ_{ij}, where ρ_{ij} is the population correlation of x_i with x_j. Thus equation (3.10) becomes

$$\lambda_p\lambda_q E(d\lambda_p d\lambda_q) = a_{ip}a_{jp}a_{kq}a_{mq}E(d\rho_{ij}d\rho_{km}), \qquad [p, q] \tag{3.22}$$

where $d\rho_{ij} = r_{ij} - \rho_{ij}$, r_{ij} being the sample correlation between x_i and x_j. The expected value of $d\rho_{ij}d\rho_{km}$ is not, as in the case of the σ's given in the simple form of (3.11) but rather it is given asymptotically, the leading term in n^{-1} being the following lengthy expression:

$$\begin{aligned} nE(d\rho_{ij}d\rho_{km}) &= \rho_{ik}\rho_{mj} + \rho_{kj}\rho_{mi} - \rho_{ij}\rho_{ki}\rho_{mi} - \rho_{ij}\rho_{kj}\rho_{mj} \\ &\quad - \rho_{km}\rho_{ki}\rho_{kj} + \tfrac{1}{2}\rho_{ij}\rho_{km}\rho_{ki}^2 + \tfrac{1}{2}\rho_{ij}\rho_{km}\rho_{kj}^2 \\ &\quad - \rho_{km}\rho_{mi}\rho_{mj} + \tfrac{1}{2}\rho_{ij}\rho_{km}\rho_{mi}^2 + \tfrac{1}{2}\rho_{ij}\rho_{km}\rho_{mj}^2 . \qquad [i, j, k, m] \end{aligned} \tag{3.23}$$

Substituting this in (3.22) and simplifying by means of equations (3.1) and (3.4) we finally get

$$\begin{aligned} n\lambda_p\lambda_q E(d\lambda_p d\lambda_q) &= 2(\lambda_p^4\delta_{pq} + \lambda_p\lambda_q a_{ip}^2 a_{jq}^2 \rho_{ij}^2) \\ &\quad - 2(\lambda_p\lambda_q^2 a_{ip}^2 a_{iq}^2 + \lambda_p^2\lambda_q a_{jp}^2 a_{jq}^2). \qquad [p, q] \end{aligned} \tag{3.24}$$

When $p = q$, (3.24) becomes

$$E[(d\lambda_p)^2] = \frac{2}{n}\left[\lambda_p^2 + a_{ip}^2 a_{jp}^2 \rho_{ij}^2 - 2\lambda_p \sum_{i=1}^{s} a_{ip}^4\right]. \qquad [p] \tag{3.25}$$

When $p \neq q$,

$$E(d\lambda_p d\lambda_q) = \frac{2}{n}[a_{ip}^2 a_{jq}^2 \rho_{ij}^2 - (\lambda_p + \lambda_q) a_{ip}^2 a_{iq}^2]. \qquad [p, q] \tag{3.26}$$

Hence (3.25) is the leading term of the asymptotic expansion of the variance of λ_p, and (3.26) is the leading term of the asymptotic expansion of the covariance of λ_p and λ_q, where λ_p and λ_q are simple, non-vanishing roots of a determinant of correlations.

4. Asymptotic Variances and Covariances of Coefficients of Principal Components Derived from a Determinant of Covariances. Let $x_i = a_{ij}\gamma_j$ be the equation of transformation of the variates $x_1, x_2, \cdots, x_s$ into principal components. In what follows we assume that the latent roots of the determinantal equation (3.3) are simple and none equal to zero. The last restriction makes the determinant of covariances non-vanishing. The determinant of the a_{ij}'s will therefore be also different from zero. With these assumptions in mind, we now proceed to derive the asymptotic variances and covariances of the a_{ij}'s.

We set $p = q$ in (3.2) and differentiate the result. This yields:

$$d\lambda_p = 2a_{lp}da_{lp}, \qquad [p] \tag{3.27}$$

where the summation index i was replaced by l. Substituting for $d\lambda_p$ from (3.8) we get:

$$a_{ip}a_{jp}d\sigma_{ij} = 2\lambda_p a_{lp}da_{lp}. \qquad [p] \tag{3.28}$$

Now, when $p \neq q$, equation (3.7) reduces to

$$\lambda_p a_{jp}da_{jq} + a_{ip}a_{jq}d\sigma_{ij} = \lambda_q a_{ip}da_{iq}, \qquad [p, q]$$

or

$$a_{ip}a_{jq}d\sigma_{ij} = (\lambda_q - \lambda_p)a_{lp}da_{lq}. \quad [p, q] \tag{3.29}$$

We combine equations (3.28) and (3.29) into one equation

$$a_{ip}a_{jq}d\sigma_{ij} = (\lambda_q + \epsilon_{pq}\lambda_p)a_{lp}da_{lq}, \quad [p, q] \tag{3.30}$$

where ϵ_{pq} has the value 1 when $p = q$ and -1 when $p \neq q$. The reciprocal of $\lambda_q + \epsilon_{pq}\lambda_p$, (which is different from zero), we denote by b_{qp}. Then equation (3.30) can be written as

$$a_{ip}b_{qp}a_{jq}d\sigma_{ij} = a_{lp}da_{lq}. \quad [p, q] \tag{3.31}$$

Since the determinant $|a_{ij}|$ of the a_{ij}'s is different from zero, we can solve this set of homogeneous linear equations for da_{lq}'s, $(l = 1, 2, \cdots, s)$. To do this we multiply equation (3.31) by A^{tp}, where A^{tp} is the element of the t^{th} row and p^{th} column of the inverse of the determinant $|a_{ij}|$, and sum with respect to p. Since $A^{tp}a_{lp} = \delta_{tl}$ we get,

$$A^{tp}a_{ip}b_{qp}a_{jq}d\sigma_{ij} = \delta_{tl}da_{lp} = da_{tq}. \quad [q] \tag{3.32}$$

We now change the subscripts i, j, t, p, q, in (3.32) to k, m, r, u, v, respectively, multiply the two equations thus obtained, and take the expected value:

$$E(da_{tq}da_{rv}) = A^{tp}A^{ru}a_{ip}a_{ku}b_{qp}b_{vu}a_{jq}a_{mv}E(d\sigma_{ij}d\sigma_{km}). \quad [q, v] \tag{3.33}$$

Substituting for $Ed\sigma_{ij}d\sigma_{km}$ its values from (3.11) and simplifying by means of (3.4) we get:

$$E(da_{tq}da_{rv}) = \frac{\lambda_v^2\lambda_q^2}{n}A^{tv}A^{rq}b_{vq}b_{qv} + \frac{\lambda_q^2\delta_{qv}}{n}\sum_{u=1}^{s}A^{tu}A^{ru}b_{qu}b_{vu}\lambda_u^2 \tag{3.34}$$

where we sum *only with respect to u*. We may simplify this formula to some extent by employing the relation: $A^{tq} = a_{tq}/\lambda_q$. (This relation is obtained from (3.2) by multiplying each side of that equation by A^{tp} and summing with respect to p). When this is done and the values for the b's are substituted, the final result becomes:

$$\begin{aligned} E(da_{tq}da_{rv}) &= \frac{\lambda_v\lambda_q a_{tv}a_{rq}}{n(\lambda_q + \epsilon_{qv}\lambda_v)(\lambda_v + \epsilon_{qv}\lambda_q)} \\ &\qquad + \frac{\lambda_q^2\delta_{qv}}{n}\sum_{u=1}^{s}\frac{a_{tu}a_{ru}}{(\lambda_q + \epsilon_{qu}\lambda_u)(\lambda_v + \epsilon_{vu}\lambda_u)}. \end{aligned} \tag{3.35}$$

From this we derive the following specific formulas:

$$\begin{aligned} E(da_{tq}da_{rq}) &= \frac{a_{tq}a_{rq}}{4n} \\ &\qquad + \frac{\lambda_q^2}{n}\left[\frac{a_{t1}a_{r1}}{(\lambda_q - \lambda_1)^2} + \cdots + \frac{a_{tq}a_{rq}}{4\lambda_q^2} + \cdots + \frac{a_{ts}a_{rs}}{(\lambda_q - \lambda_s)^2}\right] \end{aligned} \tag{3.36}$$

$$(3.37)\qquad E[(da_{tq})^2] = \frac{a_{tq}^2}{4n} + \frac{\lambda_q^2}{n}\left[\frac{a_{t1}^2}{(\lambda_q - \lambda_1)^2} + \cdots + \frac{a_{tq}^2}{4\lambda_q^2} + \cdots + \frac{a_{ts}^2}{(\lambda_q - \lambda_s)^2}\right]$$

$$(3.38)\qquad E(da_{tq}da_{rv}) = -\frac{\lambda_q\lambda_v a_{tv}a_{rq}}{n(\lambda_q - \lambda_v)^2}. \qquad (q \neq v)$$

Formulas (3.36), (3.37) and (3.38) give us the leading terms of the asymptotic expansions of the variances and covariances for the principal components. It should be remarked that the coefficients of "mutual regression" equations can be easily shown to be proportional to those of the principal components. Hence their asymptotic standard errors and covariances may be derived in a similar manner and will be of the same form.

5. **Variances and Covariances of Latent Roots when the Population Roots are Equal.** Let $k_1, k_2, \cdots, k_p$ be the latent roots of a generalized sample variance of p normally distributed variates.

Ordinarily the subscripts of the roots designate their ranks, so that $k_1 \geq k_2 \geq \cdots \geq k_p$. We may, however, assign to a root a subscript from 1 to p without any regard to its size.[11] If this is done randomly for every sample of n observations the mathematical expectation of $k_i^r k_j^s k_k^t \cdots$ will be the same for every permutation of the subscripts $i, j, k, \cdots$. This fact permits us to calculate the variances and covariances of the above roots.

We may assume, without any loss of generality, that the p variates are independently distributed,[12] and furthermore we assume the population roots to be all equal to unity. Then equation (3.11) becomes

$$(3.39)\qquad E(s_{ij}s_{km}) = \delta_{ij}\delta_{km} + \frac{1}{n}(\delta_{ik}\delta_{jm} + \delta_{im}\delta_{jk}).$$

Where s_{pq} is the sample variance of x_p and x_q and δ_{pq} is the Kronecker delta.

Now it can be easily shown that

$$(3.40)\qquad \sum_1^p s_{ii} = \sum_1^p k, \quad \sum_{i<j}(s_{ii}s_{jj} - s_{ij}^2) = \sum_{i<j} k_i k_j, \quad \sum_1^p s_{ii}^2 + 2\sum_{i<j} s_{ij}^2 = \sum_1^p k^2.$$

Hence $E(k) = 1$, and

$$E(\sum k^2) = E(\sum s_{ii}^2 + 2\sum_{i<j} s_{ij}^2)$$

or

$$(3.41)\qquad pEk^2 = pEs_{ii}^2 + p(p-1)Es_{ij}^2. \qquad (i \neq j)$$

Substituting from (3.39) in (3.41) we get

$$E(k^2) = 1 + \frac{p+1}{n}.$$

[11] This approach was suggested to the author by Professor Hotelling.

[12] See Part II, last Paragraph.

The variance of k is therefore given exactly by

$$\sigma_k^2 = E(k^2) - 1 = \frac{p+1}{n}. \tag{3.42}$$

In a similar manner we find the covariances of k_i and k_j to be

$$\sigma_{k_i k_j} = -\frac{1}{n} \tag{3.43}$$

IV. DISTRIBUTION AND MOMENTS OF QUANTITIES RELATED TO q AND z

From the known distribution of q and z and their expressions in terms of the ratio of determinants given by (1.1) and (1.12), we can derive moments and distributions of several related functions of sample variances and correlations of two independent sets of variates.

$$\text{Let } p = \frac{q^2}{z} = \frac{|b_{ij}|}{|c_{ij}|} \quad \text{by (1.12).} \tag{4.1}$$

Since the two determinants in (4.1) are independently distributed, the sampling distribution of p, given in the above form, can be obtained for a general value of s and t from Wilks'[13] distribution of the ratio of independent generalized variances.

Thus, for $s = 2$ and $t \geq 2$, the distribution of p is given by

$$\frac{\Gamma(n-2)}{2\Gamma(t-1)\Gamma(n-t-1)} p^{\frac{1}{2}(t-3)} \frac{dp}{(1+\sqrt{p})^{n-2}}. \tag{4.2}$$

When the number of variates in each set is the same, the numerator of q^2 in (1.1) becomes the square of the determinant of covariances *between* the two sets of variates. Thus

$$q^2 = \frac{|a_{i\alpha}|^2}{|a_{ij}|\,|a_{\alpha\beta}|} \tag{4.3}$$

where i, j, take on values from 1 to s, α, β take on values from $s+1$ to $2s$, and $a_{uv} = \sum_1^n x_u x_v$.

If the two sets are independent, the quantities q^2, $|a_{ij}|$, $|a_{\alpha\beta}|$, are independently distributed. Hence

$$E(|a_{i\alpha}|^m) = Eq^m(|a_{ij}|^{\frac{1}{2}m})\,E(|a_{\alpha\beta}|^{\frac{1}{2}m}). \tag{4.4}$$

Setting $\beta = 0$ in (1.16) and employing formula (1.15) we get for the moment of $|a_{i\alpha}|$

$$E(|a_{i\alpha}|^m) = \frac{2^{sm}}{|A_{ij}|^{\frac{1}{2}m}|A_{\alpha\beta}|^{\frac{1}{2}m}} \prod_{i=1}^{s} \left[\frac{\Gamma\left(\frac{s+m-i+1}{2}\right)\Gamma\left(\frac{n+m-i+1}{2}\right)}{\Gamma\left(\frac{s-i+1}{2}\right)\Gamma\left(\frac{n-i+1}{2}\right)} \right] \tag{4.5}$$

[13] Loc. cit., pp. 478–479.

where A_{uv} denotes the cofactor corresponding to σ_{uv} divided by the determinant $|\sigma_{uv}|$, σ_{uv} being the population covariance of x_u and x_v.

We may replace the product sums in (4.3) by sample correlations and, with the assumption that all the variates come from independent populations, obtain the m^{th} moment of the determinant of correlations between the two sets as

$$(4.6)\qquad E(|r_{i\alpha}|^m) = \frac{\Gamma^{2s}\left(\frac{n}{2}\right)}{\Gamma^{2s}\left(\frac{n+m}{2}\right)} \prod_{i=1}^{s} \left[\frac{\Gamma\left(\frac{n+m-i+1}{2}\right)\Gamma\left(\frac{s+m-i+1}{2}\right)}{\Gamma\left(\frac{n-i+1}{2}\right)\Gamma\left(\frac{s-i+1}{2}\right)}\right].$$

This follows from the expression for the m^{th} moment of q and the formula

$$(4.7)\qquad E(|r_{uv}|^k) = \frac{\Gamma^{s}\left(\frac{n}{2}\right)}{\Gamma^{s}\left(\frac{n+2k}{2}\right)} \prod_{i=1}^{s} \left[\frac{\Gamma\left(\frac{n+2k-i+1}{2}\right)}{\Gamma\left(\frac{n-i+1}{2}\right)}\right]$$

derived by Wilks.[14]

If we set $s = t = 2$, the numerator of q^2 in (4.3) becomes the square of a determinant of sample covariances (or correlations) known to psychologists as the tetrad. We shall here derive its distribution under the assumptions that the four variates are independently distributed.

We write

$$(4.8)\qquad q = \frac{T}{u_1 u_2}$$

where

$$(4.9)\qquad T = r_{13}r_{24} - r_{14}r_{23}, \qquad u_1 = (1 - r_{12}^2)^{\frac{1}{2}}, \qquad u_2 = (1 - r_{34}^2)^{\frac{1}{2}}$$

and q is taken as positive.

Now the distribution of q for $s = t = 2$ is given by

$$(4.10)\qquad (n-2)(1-q)^{n-3}dq$$

and the distribution of u is known to be

$$(4.11)\qquad \frac{2\Gamma\left(\frac{n}{2}\right)}{\sqrt{\pi}\,\Gamma\left(\frac{n-1}{2}\right)} u^{n-2}(1-u^2)^{-\frac{1}{2}}\,du.$$

Hence the distribution of u_1, u_2 and q is given by

$$(4.12)\qquad \frac{4(n-2)}{\pi} \frac{\Gamma^2\left(\frac{n}{2}\right)}{\Gamma^2\left(\frac{n-1}{2}\right)} (1-q)^{n-3}(u_1u_2)^{n-2}[(1-u_1^2)(1-u_2^2)]^{-\frac{1}{2}}\,du_1\,du_2\,dq.$$

[14] Loc. cit., p. 492.

Performing the transformation (4.8) and integrating out u_1 and u_2 we get for the distribution of the tetrad

$$(4.13)\qquad \frac{4(n-2)\Gamma^2\left(\frac{n}{2}\right)}{\pi\Gamma^2\left(\frac{n-1}{2}\right)}\int_T^1\int_{\frac{T}{u_1}}^1 \frac{(u_1u_2-T)^{n-3}}{\sqrt{(1-u_1^2)(1-u_2^2)}}\,du_1,\,du_2.$$

All the moments of T can of course be obtained by setting $s = 2$ in (4.6).[15]

U. S. Department of Agriculture,
Washington, D. C.

---

[15] The limiting distribution of the tetrad was given by J. L. Doob in an article entitled "The Limiting Distributions of Certain Statistics," *Annals of Mathematical Statistics*, Vol. 6, (1935). For a more general distribution of the tetrad and other statistics considered in this paper see W. G. Madow, "Contributions to the Theory of Multivariate Statistical Analysis," *Transactions of the American Mathematical Society*, Nov. 1938.

AN OPTIMUM PROPERTY OF CONFIDENCE REGIONS ASSOCIATED WITH THE LIKELIHOOD FUNCTION[1]

By S. S. Wilks and J. F. Daly

One of the authors [1] has recently established a connection between the method of maximum likelihood and shortest average confidence intervals for the case of one unknown parameter, and has reported a generalization [2] of this result for the case of several parameters. It is the object of this paper to consider the several-parameter problem in greater detail and at the same time to make the previously obtained result slightly stronger, particularly in the one-parameter case.

Let x denote a set of random variables, and θ a set of parameters $\theta_1, \cdots, \theta_h$. Suppose Π_0 is a population with the cumulative distribution function $F(x, \theta_0) \equiv F_0$ say. Then the logarithm of the likelihood associated with the population Π_0^n of random samples $0_n: x_1, x_2, \cdots, x_n$ drawn from Π_0 is

$$L^n(x, \theta_0) = \sum_{\alpha=1}^{n} \log dF(x_\alpha, \theta_0).$$

For a given sample 0_n we shall say that a set of functions $H_i^n(x, \theta)$ is of class K if there exists a domain R of parameter points θ: $(\theta_1, \cdots, \theta_h)$ in a θ-space such that for each θ_0 in R:

(i) $H_i^n(x, \theta_0) = H_{i0}^n$ is of the form $\sum_{\alpha=1}^{n} h_i(x_\alpha, \theta_0)$;
(ii) $h_i(x, \theta_0) = h_{i0}$ exists for all x except possibly for a set of zero probability;
(iii) $E_0[h_{i0}] = 0$, where E_0 means that the expected value is taken for the population Π_0;
(iv) $\| E_0[h_{i0}h_{j0}] \|$ exists and is non-singular;
(v) the moments $E_0[h_{i0}h_{j0}h_{k0}]$ are all finite.

(Here and throughout the remainder of the paper, the indices i, j, k, l have the range $1, \cdots, h$.) If, in addition,
(iii′) $E_0[h_{i0}]$ can be differentiated under the integral sign;
(iv′) the moments $E_0[h_{i0}h_{j0}]$ are differentiable with respect to the θ's;
the H_i will be said to be of class K'.

We shall need the following lemma, which is very closely related to Theorem 1′ and Theorem 2 in [1] and which can be proved by the method of characteristic functions.

---

[1] Incorporated in this paper is a note presented by one of us (c.f. [2]) at a meeting of the Institute of Mathematical Statistics, December 27, 1938.

LEMMA: *Let $H_i^n(x, \theta)$ be of class K for each n, and put*

$$B_{ij0}^n = \frac{1}{n} E_0[H_{i0}^n H_{j0}^n] = E_0[h_{i0} h_{j0}].$$

Let $\| b_{ij0}^n \|$ be the positive definite matrix satisfying the equation

$$\| b_{ij0}^n \|^2 = \| B_{ij0}^n \|$$

and write

$$\| b_{ij0}^n \|^{-1} = \| b_0^{nij} \|$$

Then for any point θ_0 in R the functions

$$\varphi_{i0}^n = \frac{1}{\sqrt{n}} \sum_{j=1}^{h} b_0^{nij} H_{j0}^n \tag{1}$$

computed from Π_0^n have a joint distribution which converges in large samples to normality, with the density function

$$(2\pi)^{-\frac{h}{2}} e^{-\frac{1}{2} \sum_{i=1}^{h} \varphi_{i0}^2}$$

Now whenever we are justified in assuming a definite functional form for $F(x, \theta)$, and have a set of functions $\varphi_i(x, \theta)$ whose distribution under this last assumption is known and is independent of the θ's, as is the case in the limit for the functions (1), we can obtain, from a sample, information about the values of the θ's. For, given any region S in the space of the functions φ_i, we can determine the probability $P_0\{\varphi_{i0} \subset S\}$ that in samples from Π_0 the point $(\varphi_{10}, \cdots, \varphi_{h0})$ will fall in the region S, even though we do not know the population values θ_0. Suppose, then, that we pick a region S such that $P_0\{\varphi_{i0} \subset S\} > .95$, and agree that each time we encounter such a problem we shall substitute the observed x's into the φ's, and call the set of all points $(\theta_1, \cdots, \theta_h)$ for which $\varphi_i(x, \theta) \subset S$ the *confidence region* T. If this procedure is followed consistently, we can assert that the probability is more than .95 that the region T thus determined contains the true parameter point θ_0.

Evidently the size of the confidence region, i.e., the accuracy with which it serves to locate the true parameter point θ_0, depends upon our choice of the auxiliary functions φ_i. Consider now the case in which there is but one parameter θ, and let $\varphi(x, \theta)$ and $\varphi^*(x, \theta)$ be two functions with the same distribution $D(u)$, where $D(u)$ does not depend on θ. For the set S of the above discussion take the interval $\underline{u} < u < \bar{u}$. Then

$$P_0\{\varphi_0 \subset S\} = P_0\{\varphi_0^* \subset S\} = \alpha$$

where $\alpha = .95$, say. Given a set of observed x's, $\varphi(x, \theta)$ will map S into a confidence region T, while $\varphi^*(x, \theta)$ will map it into a confidence region T^*. Both T and T^* may be expected to contain the true value θ_0 in 95% of the cases; hence a reasonable way to compare the size of T with that of T^* is to compare the

quantities $\frac{\partial\varphi}{\partial\theta}(x, \theta_0)$ and $\frac{\partial\varphi^*}{\partial\theta}(x, \theta_0)$; for these derivatives give an indication of the amount of change one can make in θ without forcing φ or φ^* out of the interval S.

The result obtained in [1] in this connection may now be stated as follows: Let $H = \frac{\partial L}{\partial\theta}$ be of class K', and let $H^* = \sum_{\alpha=1}^{n} h(x_\alpha, \theta)$ be any other function of class K'. Then in large samples from Π_0 both

$$\varphi = \frac{H}{\left(nE\left[\left\{\frac{\partial}{\partial\theta}\log dF\right\}^2\right]\right)^{\frac{1}{2}}}$$

and

$$\varphi^* = \frac{H^*}{(nE[\{h(x, \theta)\}])^{\frac{1}{2}}}$$

are distributed almost normally with zero mean and unit variance. But the confidence regions obtained from φ will, on the average, be smaller than those from φ^*, in the sense that, for large samples the inequality

$$\left\{E_0\left[\frac{\partial\varphi_0}{\partial\theta}\right]\right\}^2 > \left\{E_0\left[\frac{\partial\varphi_0^*}{\partial\theta}\right]\right\}^2 \tag{2}$$

will hold (unless $h(x, \theta) \equiv c\,\frac{\partial}{\partial\theta}\log dF$, in which case alone the inequality (2) becomes an equality).

Now let us return to the several-parameter case. One method of attack which suggests itself is to consider the jacobian determinant

$$\left|\frac{\partial\varphi_{i0}}{\partial\theta_j}\right|$$

for this bears the same relation to the area of the region dS which maps into the region

$$dT\colon\ \theta_0 - \tfrac{1}{2}d\theta < \theta < \theta_0 + \tfrac{1}{2}d\theta$$

as does the derivative $\frac{\partial\varphi_0}{\partial\theta}$ in the one parameter case. To this end, let us put $L_i^n(x, \theta) = \frac{\partial L^n}{\partial\theta_i}$, and for each n and for each θ_0 in R assume that

(a) L_{i0}^n is defined for all x except perhaps on a set of probability 0;
(b) $E_0[L_{i0}^n] = 0$;
(c) $E_0[L_{i0}^n]$ can be differentiated under the integral sign;
(d) $\| E_0[L_{i0}^n L_{j0}^n] \|$ exists and is non-singular;
(e) $E_0[L_{i0}^n L_{j0}^n]$ is differentiable in the θ's.

Let $H_i^n(x, \theta)$ be any other set of functions satisfying the same conditions. Set

$$E_0[L_{i0}^n L_{j0}^n] = nA_{ij0}^n \qquad E_0[H_{i0}^n H_{j0}^n] = nB_{ij0}^n$$

and define the matrices

$$\| a_{ij0}^n \|^2 = \| A_{ij0}^n \| \qquad \| a_0^{nij} \| = \| a_{ij0}^n \|^{-1}$$

$$\| b_{ij0}^n \|^2 = \| B_{ij0}^n \| \qquad \| b_0^{nij} \| = \| b_{ij0}^n \|^{-1}$$

Now consider the normalized functions

$$\tilde{L}_{i0}^n = \sum_{j=1}^h a_0^{nij} L_{j0}^n$$

$$\tilde{H}_{i0}^n = \sum_{j=1}^h b_0^{nij} H_{j0}^n$$

We then have

$$\frac{1}{n} \frac{\partial \tilde{L}_{i0}^n}{\partial \theta_k} = \sum_{j=1}^h \frac{\partial a_0^{nij}}{\partial \theta_k} \cdot \frac{1}{n} \cdot L_{j0}^n + \sum_{j=1}^h a_0^{nij} \cdot \frac{1}{n} \frac{\partial L_{j0}^n}{\partial \theta_k} \tag{3}$$

and by virtue of assumptions (b) and (c) it follows that (c.f. [1], pp. 171–2)

$$E_0\left[\frac{1}{n} \frac{\partial \tilde{L}_{i0}^n}{\partial \theta_k}\right] = -\frac{1}{n} \sum_{j=1}^h a_0^{nij} E_0[L_{j0}^n L_{k0}^n]$$

In similar fashion

$$E_0\left[\frac{1}{n} \frac{\partial \tilde{H}_{i0}^n}{\partial \theta_k}\right] = -\frac{1}{n} \sum_{j=1}^h b_0^{nij} E_0[H_{j0}^n L_{k0}^n]$$

Consequently

$$(-1)^h \left| E_0\left[\frac{1}{n} \frac{\partial \tilde{L}_{i0}^n}{\partial \theta_k}\right] \right| = | A_{ij0}^n |^{\frac{1}{2}} \tag{4}$$

and

$$(-1)^h \left| E_0\left[\frac{1}{n} \frac{\partial \tilde{H}_{i0}^n}{\partial \theta_k}\right] \right| = | B_{ij0}^n |^{-\frac{1}{2}} \cdot \left| \frac{1}{n} E_0[H_{i0}^n L_{j0}^n] \right| \tag{5}$$

We can find a relation between these two determinants by going over to the matrix

$$M_n = \left\| \begin{matrix} \| E_0[L_{i0}^n L_{j0}^n] \| & \| E_0[L_{i0}^n H_{j0}^n] \| \\ \| E_0[H_{i0}^n L_{j0}^n] \| & \| E_0[H_{i0}^n H_{j0}^n] \| \end{matrix} \right\|$$

This matrix is positive definite unless there is a linear relation with constant coefficients, say $\sum (c_i L_i + d_i H_i) = 0$, which holds for all x's except a set of zero probability; and in this event it is positive semidefinite. From the theory of compound matrices [3] we can then conclude that the matrix whose elements are the h-th order minors of M_n arranged in lexicographic order on both row and column indices has the same property, so that

$$| E_0[L_{i0}^n L_{j0}^n] | \cdot | E_0[H_{i0}^n H_{j0}^n] | \geq | E_0[L_{i0}^n H_{j0}^n] |^2$$

The relations (4) and (5) then imply that

$$\left|\det E_0\left[\frac{1}{n}\frac{\partial \tilde{L}^n_{i0}}{\partial \theta_k}\right]\right| \geq \left|\det E_0\left[\frac{1}{n}\frac{\partial \tilde{H}^n_{i0}}{\partial \theta_k}\right]\right| \tag{6}$$

It may be observed that no use has been made of the assumption of linearity (i) in deriving (6). And since in the one parameter case the determinants have but one row and column, we see that in this case the result in [1] remains valid for functions of an even more general type than those of class K'. In order to give the inequality a statistical meaning it seems necessary, however, to require not only that H and L satisfy (a), $\cdots$ (e) but also that in large samples $\frac{1}{\sqrt{n}}\tilde{H}^n_1$ and $\frac{1}{\sqrt{n}}\tilde{L}^n_1$ tend to be distributed independently of θ, with the same (though not necessarily normal) distribution.

For the case of several parameters the transition from the above determinants of expected values to the jacobian determinants requires further argument and further assumptions. To begin with, suppose that the L^n_i and H^n_i are of class K', and that

(vi) the moments $E_0\left[\frac{\partial h_{i0}}{\partial \theta_j}\frac{\partial h_{k0}}{\partial \theta_l}\right]$ are all finite,

with a corresponding condition on the variances and covariances of $\frac{\partial^2}{\partial \theta_i \partial \theta_j}\log dF(x, \theta_0)$. Let us put

$$Y^n_{ij0} = \frac{1}{n}\frac{\partial H^n_{i0}}{\partial \theta_j} - E_0\left[\frac{1}{n}\frac{\partial H^n_{i0}}{\partial \theta_j}\right]$$

$$y_{ij0} = \frac{\partial h_{i0}}{\partial \theta_j} - E_0\left[\frac{\partial h_{i0}}{\partial \theta_j}\right]$$

The characteristic function of the Y^n_{ij} is

$$\begin{aligned}\varphi_n(t_{11}, \cdots, t_{hh}) = \varphi_n(t) &= E_0[\exp(i \textstyle\sum t_{ij} Y_{ij})]\\ &= \left\{E_0\left[\exp\left(\frac{i}{n}\textstyle\sum t_{ij}y_{ij}\right)\right]\right\}^n\end{aligned}$$

Expanding the exponential in powers of the t's and using (vi), we find that

$$\varphi_n(t) = \left\{1 - O\left(\frac{1}{n^2}\right)\right\}^n$$

so that we have

$$\lim_{n\to\infty}\varphi_n(t) = 1$$

uniformly in every finite interval $|t_{ij}| < M$. A basic theorem on sequences of characteristic functions [4] then guarantees that for any $\epsilon > 0$

$$\lim_{n\to\infty} P_0\left\{\left|\frac{1}{n}\frac{\partial H^n_{i0}}{\partial \theta_j} - E_0\left[\frac{1}{n}\frac{\partial H^n_{i0}}{\partial \theta_j}\right]\right| > \epsilon\right\} = 0$$

that is to say, that $\frac{1}{n}\frac{\partial H_{i0}^n}{\partial\theta_j}$ converges stochastically to its expected value. Under the assumptions of this paragraph the same type of reasoning may be used to show that the quantities $\frac{1}{n}H_{i0}^n$, $\frac{1}{n}L_{i0}^n$, and $\frac{1}{n}\frac{\partial L_{i0}^n}{\partial\theta_j}$ all converge stochastically to their respective mean values. It will then follow from equation (3) that the functions $\frac{1}{n}\frac{\partial \tilde{L}_{i0}^n}{\partial\theta_j}$ converge stochastically to the values $E_0\left[\frac{1}{n}\frac{\partial \tilde{L}_{i0}^n}{\partial\theta_j}\right]$. In fact, it can be shown [5] that any polynomial in these functions must converge stochastically to the same polynomial in their expected values. Hence, given any $\epsilon > 0$, the probability that the determinant $\left|\frac{1}{n}\frac{\partial \tilde{L}_{i0}^n}{\partial\theta_j}\right|$ differs in samples from Π_0 from the determinant $\left|E_0\left[\frac{1}{n}\frac{\partial \tilde{L}_{i0}^n}{\partial\theta_j}\right]\right|$ by more than ϵ can be made arbitrarily small by taking n sufficiently large. Similarly, the determinant $\left|\frac{1}{n}\frac{\partial \tilde{H}_{i0}^n}{\partial\theta_j}\right|$ converges stochastically to $\left|E_0\left[\frac{1}{n}\frac{\partial \tilde{H}_{i0}^n}{\partial\theta_j}\right]\right|$. Thus, given any two positive numbers ϵ, ϵ', we have the relation

$$P_0\left\{\left|\frac{1}{n}\frac{\partial \tilde{L}_{i0}^n}{\partial\theta_j}\right|^+ > \left|\frac{1}{n}\frac{\partial \tilde{H}_{i0}^n}{\partial\theta_j}\right|^+ - \epsilon\right\} > 1 - \epsilon'$$

(where $+$ indicates the absolute values of the determinants), provided n is sufficiently large.

As in the one parameter case, the restrictions which have been put on the class of functions L and H are not entirely necessary. But it is difficult to replace them by any other set of conditions which are not obviously *ad hoc*. Let us now summarize the above results.

THEOREM 1. *If the functions L_i^n and H_i^n satisfy the conditions* (a), $\cdots$ (e), *and if*

(f) *the functions $\frac{1}{n}\frac{\partial \tilde{L}_{i0}^n}{\partial\theta_j}$ and $\frac{1}{n}\frac{\partial \tilde{H}_{i0}^n}{\partial\theta_j}$ converge stochastically to their mean values;*

(g) *the large sample distribution of the functions $\frac{1}{\sqrt{n}}\tilde{L}_{i0}^n$ is the same as that of the functions $\frac{1}{\sqrt{n}}\tilde{H}_{i0}^n$ and is independent of the θ_0's;*

then in large samples the confidence regions derived from the $\tilde{L}$'s will almost certainly be smaller than those derived from the $\tilde{H}$'s, in the sense that

$$\lim_{n\to\infty} P_0\left\{\left|\frac{1}{n}\frac{\partial \tilde{L}_{i0}^n}{\partial\theta_j}\right|^+ > \left|\frac{1}{n}\frac{\partial \tilde{H}_{i0}^n}{\partial\theta_j}\right|^+\right\} = 1$$

unless there is linear dependence between the L's and H's.

THEOREM 2. *The assumptions of Theorem 1 will be satisfied if the L_i and H_i are of class K', are linearly independent, and satisfy vi).*

THEOREM 3. *For the case of only one unknown parameter, the relation*

$$\left\{E_0\left[\frac{\partial \tilde{L}^n_{10}}{\partial \theta_1}\right]\right\}^2 \geq \left\{E_0\left[\frac{\partial \tilde{H}^n_{10}}{\partial \theta_1}\right]\right\}^2$$

$\left(\text{\textit{equality holding only in case }} H^n_1 \equiv c\,\dfrac{\partial L^n}{\partial \theta_1}\right)$ *can be derived under assumptions* (a), ..., (e) *alone. Its interpretation in terms of smallest average confidence intervals depends, however, on whether or not* (g) *is satisfied.*

At first sight it may appear that the functions

$$\psi_{ni} = \frac{1}{\sqrt{n}} \sum_{j=1}^{h} b^{nij} H^n_j$$

to which these theorems apply are too complicated to be of any practical use, involving as they do the square root of the inverse of the matrix

$$\| B^n_{ij} \| = \frac{1}{n} \| E[H^n_i H^n_j] \|.$$

But in employing the method of fiducial argument in the several parameter case there is no need to take the region S in the ψ space to be an interval

$$\underline{\psi}_i < \psi_i < \bar{\psi}_i .$$

Instead, we may take S to be the interior of the sphere

$$\sum_{i=1}^{h} \psi_i^2 < R^2 \tag{7}$$

This enables us to avoid the computation of the b^{nij}; for

$$\sum_{i=1}^{h} \psi_{ni}^2 = \frac{1}{n} \sum_{i,j,k=1}^{h} b^{nij} b^{nik} H^n_j H^n_k = \frac{1}{n} \sum_{j,k=1}^{h} B^{njk} H^n_j H^n_k$$

where $\| B^{njk} \|$ is the inverse of $\| B^n_{jk} \|$.

To indicate more precisely how the function $\sum_{i=1}^{h} \psi_{ni}^2$ may be used to determine confidence regions for the parameter point θ, we note that if the distribution law of the ψ_{ni} tends to the form

$$(2\pi)^{-\frac{h}{2}} e^{-\frac{1}{2}\Sigma \psi_i^2}$$

then $\sum_{i=1}^{h} \psi_{ni}^2$, which is identically equal to $\frac{1}{n} \sum_{i,j} B^{nij} H^n_i H^n_j$, is approximately distributed according to the χ^2 law with h degrees of freedom. We then have

$$P\left(\frac{1}{n} \sum_{i,j} B^{nij} H^n_i H^n_j < \chi^2_\alpha\right) = \alpha \tag{8}$$

approximately, where χ_α is given by the relation

$$\frac{1}{2\Gamma(\frac{1}{2}h)}\int_0^{\chi_\alpha}(\tfrac{1}{2}\chi^2)^{\frac{1}{2}h-1}e^{-\frac{1}{2}\chi^2}\,d\chi^2 = \alpha.$$

The confidence region T corresponding to a particular sample $0_n: x_1, x_2, \cdots, x_n$ consists of those points in the θ space for which $\frac{1}{n}\sum_{i,j} B^{nij}H_i^nH_j^n < \chi_\alpha^2$ when the x's are substituted in the H's. Since the region T depends on the sample, it is essentially a random variable and the probability is α that T will include the point θ_0, that is, the point in the θ-space corresponding to the values of the θ's in the population.

For example, suppose the population Π is known to have the multinomial distribution law

$$f(x_0, \cdots, x_h; p_0, \cdots, p_h) = p_0^{x_0}\cdots p_h^{x_h}$$

In this case each x has but two possible values, 0 and 1, and

$$x_0 + \cdots + x_h = 1, \qquad p_0 + \cdots + p_h = 1. \tag{9}$$

The likelihood function for random samples 0_n drawn from Π has for its logarithm

$$L^n = \sum_{\nu=0}^{h} n_\nu \log p_\nu$$

where $n_\nu = \sum_{\alpha=1}^{n} x_{\nu\alpha}$, $x_{\nu\alpha}$ being the value of x_ν for the α-th observation. Because of (9) there are only h independent parameters, say p_i $(i = 1, \cdots, h)$. Thus

$$L_i^n = \frac{n_i}{p_i} - \frac{n_0}{p_0}$$

and

$$A_{ij}^n = \frac{\delta_{ij}}{p_i} + \frac{1}{p_0}$$

where δ_{ij} is unity if $i = j$ and 0 if $i \neq j$. It is not necessary to compute the a^{nij}, for, as we have seen,

$$\sum_{i=1}^{h}(\psi_i^n)^2 = \frac{1}{n}\sum_{i,j=1}^{n} A^{nij}L_i^nL_j^n$$

And one can immediately verify that

$$A^{nij} = \delta_{ij}p_i - p_ip_j$$

so that we have

$$\sum_{i=1}^{h}\psi_{ni}^2 = \frac{1}{n}\sum_{i,j=1}^{h}(\delta_{ij}p_i - p_ip_j)\left(\frac{n_i}{p_i} - \frac{n_0}{p_0}\right)\left(\frac{n_j}{p_j} - \frac{n_0}{p_0}\right) \tag{10}$$

Since in this case the L_i^n satisfy the conditions of the lemma, we know that $\sum_{i=1}^{h} \psi_{ni}^2$ is distributed, in large samples, approximately like χ^2 with h degrees of freedom.

As a matter of fact, (10) is precisely the Pearson χ^2 which is ordinarily used, in connection with the problem of deciding whether a sample supports the hypothesis that the population from which it has been drawn has specified values of the p's. For, making use of the fact that

$$\sum_{i=1}^{h} (n_i - np_i) + (n_0 - np_0) = 0$$

we find that

$$\frac{n_i}{p_i} - \frac{n_0}{p_0} = \sum_{j=1}^{h} A_{ij}^n (n_j - np_j)$$

so that $\sum_{i=1}^{h} \psi_{ni}^2$ reduces to

$$\frac{1}{n} \sum_{i,j=1}^{h} A_{ij}^n (n_i - np_i)(n_j - np_j) = \sum_{\nu=0}^{h} (n_\nu - np_\nu)^2/np_\nu$$

which is the familiar form. Thus in particular the Pearson χ^2 is the best fiducial function of its type which can be formed from H's satisfying Theorem 1, in the sense that for sufficiently large samples its constituent functions $\tilde{L}_i^n$ will almost certainly have a greater jacobian with respect to the parameters p_i than will the corresponding $\tilde{H}_i^n$ computed from a set of H_i^n independent of the L_i^n.

The confidence regions determined by (8) when the H_i^n are replaced by the L_i^n have an associated optimum property which may be stated as

THEOREM 4: *Let Δ_0 denote the differential of $\frac{1}{n}\sum_{i,j} B^{nij} H_i^n H_j^n$ with respect to the θ_i, evaluated at the true parameter point θ_0. Let Δ_0^* be the corresponding differential when the H_i^n are replaced by the L_i^n. Let the H_i^n and L_i^n satisfy conditions* (i), (ii), $\cdots$, (vi) *and let the mean value of the product of two, three or four factors taken from the set $\left\{h_{i0}, \frac{\partial h_{j0}}{\partial \theta_k}\right\}$ be finite, no product containing more than two factors of the type $\frac{\partial h_{i0}}{\partial \theta_j}$. Let similar assumptions hold for the set $\left\{l_{i0}, \frac{\partial l_{i0}}{\partial \theta_i}\right\}$ where $l_{i0} = \frac{\partial \log dF_0}{\partial \theta_i}$. Then if n is sufficiently large*

$$E_0\left(\frac{1}{n}\Delta_0^{*2}\right) - E_0\left(\frac{1}{n}\Delta_0^2\right) \geq 0 \tag{11}$$

The equality in (11) *will hold for all differential vectors if and only if each h_{i0} is a linear function of the l_{i0}.*

This theorem can be proved in a straightforward manner by using the following characteristic functions

$$\varphi_H = \exp\left(i \sum_{i=1}^{h} t_i H_{i0}^n + i \sum_{i,j=1}^{h} u_{ij} \frac{\partial H_{i0}^n}{\partial \theta_j}\right)$$

$$= \left[\exp\left(i \sum_{i=1}^{h} t_i h_{i0} + i \sum_{i,j=1}^{h} u_{ij} \frac{\partial h_{i0}}{\partial \theta_j}\right)\right]^n$$

$$\varphi_L = \exp\left(i \sum_{i=1}^{h} t_i L_i^n + i \sum_{i,j=1}^{h} u_{ij} \frac{\partial L_i^n}{\partial \theta_j}\right)$$

$$= \left[\exp\left(i \sum_{i=1}^{h} t_i l_{i0} + i \sum_{i,j=1}^{h} u_{ij} \frac{\partial l_{i0}}{\partial \theta_j}\right)\right]^n,$$

where $u_{ij} \equiv u_{ji}$. Now

$$\Delta_0 = \frac{1}{n} \sum_{i,j,k=1}^{h} \frac{\partial B^{nij}}{\partial \theta_k} H_i^n H_j^n \, d\theta_k + \frac{2}{n} \sum_{i,j,k=1}^{h} B^{nij} \frac{\partial H_i^n}{\partial \theta_k} H_j^n \, d\theta_k$$

with a similar expression for Δ_0^*. The problem of finding the mean values $E\left(\frac{1}{n}\Delta_0^2\right)$ and $E\left(\frac{1}{n}\Delta_0^{*2}\right)$ is a matter of evaluating a set of fourth order derivatives of φ_H and φ_L at $t_i = 0$, $u_{ij} = 0$.

If the appropriate differentiations are carried out it is found that

$$E_0(\Delta_0^2) = 4n\left[\sum_{i,j,k,l} B_{kl0} C_{ki0} C_{lj0} \, d\theta_i d\theta_j + 0\left(\frac{1}{n}\right)\right]$$

$$E_0(\Delta_0^{*2}) = 4n\left[\sum_{i,j,k,l} A_{ij0} \, d\theta_i d\theta_j + 0\left(\frac{1}{n}\right)\right]$$

where $A_{ij0} = E_0[l_{i0} l_{j0}]$, $B_{kl0} = E_0[h_{k0} h_{l0}]$, $C_{ki0} = E_0[h_{k0}, l_{i0}]$. Denoting $E_0\left(\frac{1}{n}\Delta_0^{*2}\right) - E_0\left(\frac{1}{n}\Delta_0^2\right)$ by δ, we have

$$\delta = 4\left\{\sum_{i,j} M_{ij0} \, d\theta_i d\theta_j + 0\left(\frac{1}{n}\right)\right\}$$

where $\| M_{ij0} \| = \| A_{ij0} - \sum_{k,l} B_0^{kl} C_{ki0} C_{lj0} \|$. If the h_{k0} and l_{i0} are linearly independent then $\| M_{ij0} \|$ is a positive definite matrix and hence $\sum_{i,j} M_{ij0} \, d\theta_i \, d\theta_j \equiv \delta'$ say, will be non-negative and can vanish only when all $d\theta_i$ are zero. If each h_{k0} is a linear combination of the l_{i0} and if the h_{k0} are linearly independent, then each l_{i0} is a linear combination of the h_{k0}. In this case it can be readily shown that every element in $\| M_{ij0} \|$ will vanish, and hence $\delta' \equiv 0$.

In case some of the h_{i0} are linearly dependent on the l_{j0}, it can be shown that δ' is positive semidefinite, that is, there exists no differential vector for which δ' is negative, although there will exist non-zero differential vectors for which δ' is zero.

It can be shown under the assumptions made in Theorem 4 that $\frac{1}{n}(\Delta_0^{*2} - \Delta_0^2)$ actually converges stochastically to $4\delta'$, and thus if the h_{i0} and l_{i0} are linearly independent, the difference $\frac{1}{n}(\Delta_0^{*2} - \Delta_0^2)$ converges stochastically to a positive number. Stated in another way: for sufficiently large samples, the square of the differential change in $\frac{1}{n}\sum_{i,j} A^{nij}L_i^n L_j^n$, for a given change $d\theta_i$ in the θ_i from the values θ_{i0}, will almost certainly exceed that of $\frac{1}{n}\sum_{i,j} B^{nij}H_i^n H_j^n$. The statistical interpretation of this result amounts to the following: by taking sufficiently large samples, we can make it as certain as we please that the confidence regions for locating θ_0 determined by using $\frac{1}{n}\sum_{i,j} A^{nij}L_i^n L_j^n$ in (8) will be smaller than those determined by using $\frac{1}{n}\sum_{i,j} B^{nij}H_i^n H_j^n$ in (8).

REFERENCES

[1] S. S. Wilks, "Shortest Average Confidence Intervals from Large Samples," *Annals of Mathematical Statistics*, vol. IX, No. 3 (1938), pp. 166–175.

[2] S. S. Wilks, "Optimum Fiducial Regions for Simultaneous Estimation of Several Population Parameters from Large Samples," (Abstract), *Annals of Mathematical Statistics*, vol. X, No. 1 (1939), p. 85.

[3] C. C. MacDuffee, *The Theory of Matrices*, Ergebnisse Series, vol. II, No. 5, Berlin, Julius Springer (1933), p. 87.

[4] H. Cramér, *Random Variables and Probability Distributions*, Cambridge Tracts in Mathematics and Mathematical Physics, No. 36, Cambridge University Press, 1937.

[5] J. L. Doob, "Limiting Distributions of Certain Statistics," *Annals of Mathematical Statistics*, vol. VI, No. 3 (1935), p. 164.

PRINCETON UNIVERSITY, PRINCETON, N. J.

ON SOME PROPERTIES OF MULTIDIMENSIONAL DISTRIBUTIONS

By J. Lukomski

If, in a system of random variables $x_1, x_2, \cdots, x_n$, some variables are connected by a functional (exact) dependence, the n-dimensional distribution law has a degenerated character. In other words, in this case the probability is not distributed over the whole n-dimensional space, but is concentrated on a manifold of a smaller number of dimensions which may be called the *skeleton of the distribution.*

The character and the dimensionality of this manifold are determined by the character and the number of functional connections between the variables $x_1, x_2, \cdots, x_n$. If all these connections are linear, the skeleton will be a linear manifold (hyperplane). The investigation of the skeleton of distribution represents obviously an interest from the theoretical as well as from the practical point of view.

In the present paper we establish some criteria which enable us to determine, for any distribution possessing finite moments of the first and second order, the linear skeleton and to find the variations of the dimensionality of this manifold when the variables are subjected to a linear transformation.[1]

We also apply the obtained results to the case of a multidimensional normal distribution (generalized by H. Cramér to the case of linear dependence between variables).

§1

Let

$$x_1, x_2, \cdots, x_n \tag{1}$$

be a system of random variables defined in the n-dimensional euclidean space R_n by the multidimensional distribution function $F(x_1, x_2, \cdots, x_n)$. The function F is defined on all Borel sets in R_n. We assume the existence of the following moments:

$$E(x_i) = \int\int \cdots \int_{R_n} x_i dd \cdots dF(x_1, x_2, \cdots x_n) = 0$$

$$E(x_i x_j) = \int\int \cdots \int_{R_n} x_i x_j dd \cdots dF(x_1, x_2, \cdots x_n) = \mu_{ij}$$

where the integrals are to be understood in the sense of Lebesgue-Radon.

[1] The questions of degeneracy of a statistical distribution were for the first time considered—from a somewhat different point of view—by R. Frisch [1].

If the variables $x_1, x_2, \cdots, x_n$ are connected by a relation of the form $C_1x_1 + C_2x_2 + \cdots + C_nx_n = 0$ $(\Sigma C^2 \neq 0)$ (are linearly dependent), we call this relation a *linear bond of the distribution* F.

We shall call a system of linear bond of the distribution F *complete*, if all bonds of the system are linearly independent and every linear bond of the distribution depends linearly on the bonds of the system.

By the (linear) *decrement* of the distribution F (we denote it by $k(F)$ or simply k) we understand the number of bonds in a complete system. We may, correspondingly, call the difference between the number of variables and the decrement of the distribution the (linear) *rank* of the distribution, or the dimensionality of the linear skeleton.

The decrement (rank) is given by the following

THEOREM 1.[2] *The decrement (rank) of the distribution F is equal to the decrement*[3] *(rank) of the matrix*

$$\| \mu_{ij} \| \qquad i, j = 1, 2, \cdots n$$

of the moments of the second order of this distribution; that is

$$k(F) = k(\| \mu_{ij} \|), \qquad i, j = 1, 2, \cdots n \tag{2}$$

PROOF. Consider the form

$$v = t_1x_1 + t_2x_2 + \cdots + t_nx_n \tag{3}$$

where $t_1, t_2, \cdots, t_n$ are arbitrary real numbers, not all equal to zero. Let

$$\begin{aligned} Q^2 = E(v^2) &= \int\int \cdots \int_{R_n} (t_1x_1 + t_2x_2 + \cdots + t_nx_n)^2 \, dd \cdots \\ &\qquad \cdots dF(x_1, x_2, \cdots x_n) \\ &= \sum_{i,j=1}^{n} t_it_j \int\int \cdots \int_{R_n} x_ix_j \, dd \cdots dF(x_1, x_2, \cdots x_n) = \sum_{i,j=1}^{n} t_it_j\mu_{ij}. \end{aligned} \tag{4}$$

Q^2 is a non-negative quadratic form in the variables $t_1, t_2, \cdots, t_n$. The system of values $t_1, t_2, \cdots, t_n$, for which the expression (3) becomes zero is a double point of the form Q^2.

The coordinates of the double point can be found from the system of homogeneous equations:

$$\begin{aligned} \mu_{11}t_1 + \mu_{12}t_2 + \cdots + \mu_{1n}t_n &= 0 \\ \mu_{21}t_1 + \mu_{22}t_2 + \cdots + \mu_{2n}t_n &= 0 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \mu_{n1}t_1 + \mu_{n2}t_2 + \cdots + \mu_{nn}t_n &= 0. \end{aligned} \tag{5}$$

---

[2] This theorem was proved by a different method by R. Frisch [1].

[3] By the decrement of a (rectangular) matrix we call, after B. Kagan, the difference between the number of its rows and its rank.

It is, however, known that the number of the independent double points of the form, Q^2, i.e. the number of linearly independent untrivial solutions of the system (5) is equal to the decrement of the matrix $\| \mu_{ij} \|$, $i, j = 1, 2, \cdots n$.

Consequently, there exist only $k(\| \mu_{ij} \|)$ independent linear connections between the variables $x_1, x_2, \cdots, x_n$, which proves the theorem.

Hence it follows that the variables $x_1, x_2, \cdots, x_n$ are linearly independent ($k(F) = 0$) if and only if the form Q^2 is positive definite and, consequently, the discriminant $| \mu_{ij} |$ of the form is positive.

The following two theorems may be used for determination of a complete system of linear bonds. The first of them is a special case of the second, but is stated separately in order to simplify the proof.

THEOREM 2. *If $k(F) = 1$, we obtain the linear bond of the distribution by replacing in the determinant on the left hand side of the equation*

$$(6) \qquad \begin{vmatrix} \mu_{11} & \mu_{12} & \cdots & \mu_{1n} \\ \mu_{21} & \mu_{22} & \cdots & \mu_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ \mu_{n1} & \mu_{n2} & \cdots & \mu_{nn} \end{vmatrix} = 0$$

the elements of one (arbitrary) row by $x_1, x_2, \cdots, x_n$ respectively.

For instance, replacing the first row, we have

$$(7) \qquad \begin{vmatrix} x_1 & x_2 & \cdots & x_n \\ \mu_{21} & \mu_{22} & \cdots & \mu_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ \mu_{n1} & \mu_{n2} & \cdots & \mu_{nn} \end{vmatrix} = 0.$$

PROOF. Since the decrement of the matrix $\| \mu_{ij} \|$, $i, j = 1, 2, \cdots n$ is equal to 1, for the unique nontrivial independent solution of the system (5) ($t_1, t_2, \cdots, t_n$) may be taken, as we know, the system of algebraical supplements of the elements of any row of the determinant $| \mu_{ij} |$, $i, j = 1, 2, \cdots n$. (Among the algebraical supplements of elements of each row there is at least one different from zero, since the algebraical supplements of corresponding elements of any pair of rows are proportional to each other.)

Hence, since $t_1x_1 + t_2x_2 + \cdots + t_nx_n = 0$, the theorem follows.

THEOREM 3. *If $k(F) > 0$, we obtain a complete system of linear bonds of the distribution F replacing in each of the k equations*

$$(8) \qquad \begin{vmatrix} \mu_{ki} & \mu_{k,k+1} & \cdots & \mu_{kn} \\ \mu_{k+1,i} & \mu_{k+1,k+1} & \cdots & \mu_{k+1,n} \\ \cdots & \cdots & \cdots & \cdots \\ \mu_{ni} & \mu_{n,k+1} & \cdots & \mu_{nn} \end{vmatrix} = 0, \qquad i = 1, 2, \cdots k$$

one (arbitrary) row of the determinant respectively by $x_i, x_{k+1}, \cdots, x_n$, where $x_{k+1}, \cdots, x_n$ are chosen in such a way that

$$\begin{vmatrix} \mu_{k+1,k+1} & \cdots & \mu_{k+1,n} \\ \cdots & \cdots & \cdots \\ \mu_{n,k+1} & \cdots & \mu_{nn} \end{vmatrix} > 0.$$

Replacing, for example, the first rows, we obtain:

$$
(9)\qquad
\begin{aligned}
&\begin{vmatrix} x_1 & x_{k+1} & \cdots & x_n \\ \mu_{k+1,1} & \mu_{k+1,k+1} & \cdots & \mu_{k+1,n} \\ \cdots & \cdots & \cdots & \cdots \\ \mu_{n1} & \mu_{n,k+1} & \cdots & \mu_{nn} \end{vmatrix} = 0 \\
&\begin{vmatrix} x_2 & x_{k+1} & \cdots & x_n \\ \mu_{k+1,2} & \mu_{k+1,k+1} & \cdots & \mu_{k+1,n} \\ \cdots & \cdots & \cdots & \cdots \\ \mu_{n2} & \mu_{n,k+1} & \cdots & \mu_{nn} \end{vmatrix} = 0 \\
&\cdots\cdots\cdots\cdots\cdots\cdots \\
&\begin{vmatrix} x_k & x_{k+1} & \cdots & x_n \\ \mu_{k+1,k} & \mu_{k+1,k+1} & \cdots & \mu_{k+1,n} \\ \cdots & \cdots & \cdots & \cdots \\ \mu_{nk} & \mu_{n,k+1} & \cdots & \mu_{nn} \end{vmatrix} = 0.
\end{aligned}
$$

PROOF. The theorem is already proved for $k(F) = 1$ (Theorem 2). We have to prove it for $k(F) > 1$.

Let us in the first place show that the matrix $\| \mu_{ij} \|$, $i, j = 1, 2, \cdots n$ possesses at least one positive chief algebraical supplement of the order $n - k$.

In fact, in the system of n variables $x_1, x_2, \cdots, x_n$, connected by k independent linear relations there must exist a subsystem of $n - k$ linearly independent variables. Let these variables be $x_{k+1}, x_{k+2}, \cdots, x_n$. The determinant of the moments of the second order of this subsystem: $| \mu_{ij} |$, $i, j = k + 1, \cdots, n$ is different from zero and, by the property of Gramm's determinants, is positive. Further, each of the subsystems $x_i, x_{k+1}, \cdots, x_n$, is subjected to the distribution law $F_i(x_i, x_{k+1}, \cdots, x_n)$ with the decrement $k_i = 1$ and, consequently, by Theorem 2, the relations (9) are satisfied. (Arguing as before we find that any (not necessarily the first) row in each of the determinants in (8) may be replaced by $x_i, x_{k+1}, \cdots x_n$).

In order to show the independence of the relations (9), write the system (9) in the form:

$$
(9') \qquad \sum_{j=1}^{n} C_{ij} x_j = 0, \qquad i = 1, 2, \cdots, k
$$

and consider the matrix of its coefficients:

$$
(10) \qquad \begin{Vmatrix} C_{11} & 0 & \cdots & 0 & C_{1,k+1} & C_{1,k+2} & \cdots & C_{1n} \\ 0 & C_{22} & \cdots & 0 & C_{2,k+1} & C_{2,k+2} & \cdots & C_{2n} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & C_{kk} & C_{k,k+1} & C_{k,k+2} & \cdots & C_{kn} \end{Vmatrix}.
$$

The matrices (10) have the rank k, since the determinant of order k

$$
\begin{vmatrix} C_{11} & 0 & \cdots & 0 \\ 0 & C_{22} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & C_{kk} \end{vmatrix} = C_{11} \cdot C_{22} \cdots C_{kk}
$$

belonging to the matrix, is positive; this follows from

$$C_{11} = C_{22} = \cdots = C_{kk} = |\mu_{ij}| > 0, \qquad i, j = k + 1, \cdots, n.$$

Thus the independence of the relations (9) is proved and the theorem is established.

§2

In this section we consider the question of the variation of decrement of the distribution in the case when the variables are subjected to a linear transformation.

Let $x_1, x_2, \cdots, x_n$ be a system (1) of random variables and

$$\begin{aligned} u_1 &= a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ u_2 &= a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ u_m &= a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{aligned} \tag{11}$$

a system of linear forms in the variables (1).

The distribution function of the variables $u_1, u_2, \cdots, u_m$ we denote by F_1, the decrement of the distribution by $k(F_1)$, or, shorter, by k_1.

The two systems of equations (11) and (9) form together the system:

$$\begin{aligned} u_1 &= a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ u_m &= a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \\ 0 &= a_{m+1,1}x_1 + a_{m+1,2}x_2 + \cdots + a_{m+1,n}x_n \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ 0 &= a_{m+k,1}x_1 + a_{m+k,2}x_2 + \cdots + a_{m+k,n}x_n \end{aligned} \tag{12}$$

where the last k equations represent, in new notation, the equations (9).

We call the matrix of the coefficients of the variables in the system (12): $\| a_{ij} \|$ $i = 1, 2, \cdots, m + k; j = 1, 2, \cdots, n,$ *the elongated matrix of the transformation.*

We prove the following

THEOREM 4. *The decrement of the distribution $F_1(u_1, u_2, \cdots u_m)$ is equal to the decrement of the elongated matrix of the transformation.*

$$k(F_1) = k(\| a_{ij} \|). \qquad \begin{aligned} i &= 1, 2, \cdots, m + k \\ j &= 1, 2, \cdots, n \end{aligned} \tag{13}$$

PROOF. Consider a system of forms in arbitrary linearly independent parameters $\xi_1, \xi_2, \cdots, \xi_n$:

$$\begin{aligned}
v_1 &= a_{11}\xi_1 + a_{12}\xi_2 + \cdots + a_{1n}\xi_n \\
v_2 &= a_{21}\xi_1 + a_{22}\xi_2 + \cdots + a_{2n}\xi_n \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
v_m &= a_{m1}\xi_1 + a_{m2}\xi_2 + \cdots + a_{mn}\xi_n \\
v_{m+1} &= a_{m+1,1}\xi_1 + a_{m+1,2}\xi_2 + \cdots + a_{m+1,n}\xi_n \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
v_{m+k} &= a_{m+k,1}\xi_1 + a_{m+k,2}\xi_2 + \cdots + a_{m+k,n}\xi_n
\end{aligned} \tag{14}$$

such that the matrix of the system (14) coincides with the elongated matrix of the transformation.

For

$$v_{m+1} = 0, \qquad v_{m+2} = 0, \cdots, \qquad v_{m+k} = 0 \tag{15}$$

the system (14) reduces to the system (12).

If the decrement of the matrix of the system is equal to s, there exist only $m + k - s$ linearly independent forms v_i, and each of the remaining s forms is a linear combination of the first.

By Steinitz's theorem we can always include in a subsystem of independent forms the forms $v_{m+1}, \cdots, v_{m+k}$ (since these forms are independent).

Denoting all forms of the subsystem by $v_{s+1}, \cdots, v_m, v_{m+1}, \cdots, v_{m+k}$, let us write the s relations connecting each of the remaining forms with the forms of our subsystem in the form:

$$\begin{aligned}
g_{11}v_1 + g_{1,s+1}v_{s+1} + \cdots + g_{1m}v_m + g_{1,m+1}v_{m+1} + \cdots + g_{1,m+k}v_{m+k} &= 0 \\
g_{22}v_2 + g_{2,s+1}v_{s+1} + \cdots + g_{2m}v_m + g_{2,m+1}v_{m+1} + \cdots + g_{2,m+k}v_{m+k} &= 0 \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots & \\
g_{ss}v_s + g_{s,s+1}v_{s+1} + \cdots + g_{sm}v_m + g_{s,m+1}v_{m+1} + \cdots + g_{s,m+k}v_{m+k} &= 0
\end{aligned} \tag{16}$$

where $g_{11}, g_{22}, \cdots, g_{ss} \neq 0$.

Assigning to the variables in these equations the values (15) we clearly obtain s linear relations between the variables $u_1, u_2, \cdots, u_m$

$$\begin{aligned}
g_{11}u_1 + g_{1,s+1}u_{s+1} + \cdots + g_{1m}u_m &= 0 \\
g_{22}u_2 + g_{2,s+1}u_{s+1} + \cdots + g_{2m}u_m &= 0 \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots & \\
g_{ss}u_s + g_{s,s+1}u_{s+1} + \cdots + g_{sm}u_m &= 0.
\end{aligned} \tag{17}$$

The equations (17) are linearly independent, since the matrix of the system (17)

$$\left\| \begin{matrix} g_{11} & 0 & \cdots & 0 & g_{1,s+1} & \cdots & g_{1m} \\ 0 & g_{22} & \cdots & 0 & g_{2,s+1} & \cdots & g_{2m} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & g_{ss} & g_{s,s+1} & \cdots & g_{sm} \end{matrix} \right\|$$

has the rank s; this follows from the fact that it contains the determinant

$$\left| \begin{matrix} g_{11} & 0 & \cdots & 0 \\ 0 & g_{22} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & g_{ss} \end{matrix} \right| = g_{11} \cdot g_{22} \cdots g_{ss}$$

of the order s, which is different from zero.

We proceed now to prove that there exists no other linear relation between the variables u_i linearly independent of the relations (17).

From the equations (17) the variables $u_1, u_2, \cdots, u_s$ may be determined as linear combinations of the variables $u_{s+1}, \cdots, u_m$ (we suppose that $m > s$, since for $m = s$ the proposition under consideration is trivial).

It is thus to be proved that the variables $u_{s+1}, \cdots, u_m$ are linearly independent (since every new linear relation between the variables $u_1, u_2, \cdots, u_m$, independent of (17) must, after corresponding substitutions, lead to a linear relation between $u_{s+1}, \cdots, u_m$).

In the equations (12) the linearly independent variables $u_{s+1}, \cdots, u_{m+k}$ are linear forms in n linearly independent parameters $\xi_1, \xi_2, \cdots, \xi_n$.

We may instead of the $\xi_1, \xi_2, \cdots, \xi_n$ take for the system of linearly independent parameters $v_{m+1}, \cdots, v_{m+k}, \xi_{k+1}, \cdots, \xi_n$ (changing the indices of the ξ in an appropriate manner), defining $\xi_1, \xi_2, \cdots, \xi_k$[4] by the system of equations

$$\begin{aligned} v_{m+1} &= a_{m+1,1}\xi_1 + \cdots + a_{m+1,n}\xi_n \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ v_{m+k} &= a_{m+k,1}\xi_1 + \cdots + a_{m+k,n}\xi_n \end{aligned}$$

which is always possible, since the forms $v_{m+1}, \cdots, v_{m+k}$ are independent.

Substituting the expressions obtained for the $\xi_1, \xi_2, \cdots, \xi_n$ into the forms $v_{s+1}, \cdots, v_m$, we find

$$\begin{aligned} v_{s+1} &= \varphi_{s+1}(v_{m+1}, \cdots v_{m+k}) + \psi_{s+1}(\xi_{k+1}, \cdots \xi_n) \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ v_m &= \varphi_m(v_{m+1}, \cdots v_{m+k}) + \psi_m(\xi_{k+1}, \cdots \xi_n) \end{aligned} \tag{18}$$

where φ and ψ are linear forms in the corresponding arguments.

The variables $v_{s+1}, \cdots, v_m$ remain, of course, independent.

[4] The indices of the ξ adequately chosen.

Performing in the equations (18) the substitution (15), we obtain:

$$
(19)\qquad
\begin{aligned}
u_{s+1} &= \psi_{s+1}(\xi_{k+1}, \cdots \xi_n)\\
&\cdots\cdots\cdots\cdots\cdots\cdots\\
u_m &= \psi_m(\xi_{k+1}, \cdots \xi_n).
\end{aligned}
$$

If there exists a linear dependence between the $u_{s+1}, \cdots, u_m$, we can find $\alpha_{s+1}, \cdots, \alpha_m$, not all equal to zero, such that

$$
(20)\qquad \alpha_{s+1}u_{s+1} + \cdots + \alpha_m u_m = 0.
$$

Multiplying the equations (18) by the coefficients $\alpha_{s+1}, \cdots, \alpha_m$ respectively, and adding, we obtain, by virtue of (19) and (20)

$$
\alpha_{s+1}v_{s+1} + \cdots + \alpha_m v_m = \alpha_{s+1}\varphi_{s+1}(v_{m+1}, \cdots v_{m+k}) + \cdots + \alpha_m\varphi_m(v_{m+1}, \cdots v_{m+k})
$$

i.e. the variables $v_{s+1}, \cdots, v_{m+k}$ are linearly dependent, which contradicts the assumption.

The required proposition is thus proved.

It follows that the s equations (17) form a complete system of bonds of the distribution F_1, which proves our theorem.

The moments of the second order of the distribution F_1 are connected with the moments of the distribution F by the following formulae

$$
(21)\qquad
\begin{aligned}
v_{ij} &= E(u_i u_j) = E\left[\left(\sum_{r=1}^{n} a_{ir}x_r\right)\left(\sum_{s=1}^{n} a_{js}x_s\right)\right]\\
&= \sum_{r,s=1}^{n} a_{ir}a_{js}E(x_r x_s) = \sum_{r,s=1}^{n} a_{ir}a_{js}\mu_{rs} \qquad (i, j = 1, 2, \cdots m).
\end{aligned}
$$

§3

Let the normal law of distribution G (generalized by H. Cramér) be given by its multidimensional characteristic function [2], [3]:

$$
(22)\qquad
\begin{aligned}
f(t_1, t_2, \cdots t_n) &= \int\int\cdots\int_{R_n} e^{i(t_1x_1+t_2x_2+\cdots+t_nx_n)}\, d\, d \cdots dG(x_1, x_2, \cdots x_n)\\
&= e^{-\frac{1}{2}Q^2}
\end{aligned}
$$

where $Q^2 = \sum_{r,s=1}^{n} c_{rs}t_r t_s$ $(c_{rs} = c_{sr})$ is a non-negative quadratic form in the real variables $t_1, t_2, \cdots, t_n$. (The integrals, as above, to be understood in the sense of Lebesgue-Radon.)

As is easily seen, the coefficients c_{rs} are the moments of the second order of the distribution G for which

$$
\mu_{rs} \equiv E(x_r x_s) \equiv i^2\left[\frac{\partial^2 f}{\partial t_r\, \partial t_s}\right]_{\Sigma t^2=0} = c_{rs}.
$$

If Q^2 is positive definite, we have a proper normal distribution.

If Q^2 is non-negative, the distribution G possesses a positive decrement.

The decrement and the linear bonds of the distribution may be determined from the matrix of the coefficients $\| c_{rs} \|$ $r, s = 1, 2, \cdots, n$ on ground of the general theorems of §1.

Let, as before,

$$
(11) \qquad \begin{aligned}
u_1 &= a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\
u_2 &= a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
u_m &= a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n
\end{aligned}
$$

be a system of linear forms in the variables $x_1, x_2, \cdots, x_n$. We shall prove the following

THEOREM 5. *The variables $u_1, u_2, \cdots, u_m$ are subject to the generalized normal distribution law the decrement of which is equal to the decrement of the elongated matrix of the transformation*

$$
\left\| \begin{matrix}
a_{11} & a_{12} & \cdots & a_{1n} \\
\cdots & \cdots & \cdots & \cdots \\
a_{m1} & a_{m2} & \cdots & a_{mn} \\
a_{m+1,1} & a_{m+1,2} & \cdots & a_{m+1,n} \\
\cdots & \cdots & \cdots & \cdots \\
a_{m+k,1} & a_{m+k,2} & \cdots & a_{m+k,n}
\end{matrix} \right\| .
$$

PROOF. Consider the characteristic function of the distribution $G_1(u_1, u_2, \cdots u_m)$,

$$
(23) \quad f_1(t_1, t_2, \cdots t_m) = \iint \cdots \int_{R_m} e^{i(t_1u_1+t_2u_2+\cdots+t_mu_m)}\, d\, d \cdots dG_1(u_1, u_2 \cdots, u_m).
$$

Performing in this expression the substitution (11), we obtain

$$
(24) \quad \begin{aligned}
&f_1(t_1, t_2, \cdots t_m) \\
&= \iint \cdots \int_{R_n} e^{i\left(t_1 \sum_{j=1}^{n} a_{ij}x_j + t_2 \sum_{j=1}^{n} a_{ij}x_j + \cdots + t_m \sum_{j=1}^{n} a_{mj}x_j\right)}\, d\, d \cdots \\
&\qquad\qquad \cdots dG_1\left\{ \sum_{j=1}^{n} a_{1j}x_j, \sum_{j=1}^{n} a_{ij}x_j, \cdots \sum_{j=1}^{n} a_{mj}x_j \right\} \\
&= \iint \cdots \int_{R_n} e^{i\left(x_1 \sum_{p=1}^{m} a_{p1}t_p + x_2 \sum_{p=1}^{m} a_{p2}t_p + \cdots + x_n \sum_{p=1}^{m} a_{pn}t_p\right)}\, d\, d \\
&\qquad\qquad \cdots dG(x_1, x_2, \cdots x_n).
\end{aligned}
$$

($d\, d \cdots dG(x_1, x_2, \cdots x_n)$ in the expression (24) does not, in general, coincide with $d\, d \cdots dG(x_1, x_2, \cdots x_n)$ in the expression (22)).

Taking into account (22), we obtain

$$
f_1 = e^{-\frac{1}{2}Q_1^2}
$$

where

$$
\begin{aligned}
Q_1^2 &= \sum_{r,s=1}^{n}\left\{c_{rs}\left(\sum_{p=1}^{m} a_{pr}t_p\right)\left(\sum_{q=1}^{m} a_{qs}t_q\right)\right\} \\
&= \sum_{r,s=1}^{n}\left\{c_{rs}\sum_{p,q=1}^{m} a_{pr}q_{qs}t_pt_q\right\} \qquad (26)\\
&= \sum_{p,q=1}^{m}\left\{t_pt_q\sum_{r,s=1}^{n} a_{pr}a_{qs}c_{rs}\right\} = \sum_{p,q=1}^{m} t_pt_qv_{pq}.
\end{aligned}
$$

Q_1^2 is a non-negative quadratic form in $t_1, t_2, \cdots, t_m$, the coefficients of which coincide with the moments of the second order of the distribution $G_1(u_1, u_2, \cdots u_m)$.

Consequently, the distribution G_1 is a generalized normal distribution.

By Theorem 4 the decrement of the distribution G_1 is equal to the decrement of the matrix $\| a_{pr} \|$ $p = 1, 2, \cdots, m + k$; $r = 1, 2, \cdots, n$, the last k rows of which consist of the coefficients of the complete system of linear bonds of the distribution G.

Let now $x_1, x_2, \cdots, x_n$ be a system of random variables subjected to a proper Gaussian law. The density function of the distribution of the system is

$$
y = Ce^{-\chi^2} = \frac{1}{(2\pi)^{\frac{n}{2}}\prod_{i=1}^{n}\sqrt{\mu_{ii}}\sqrt{R}}\exp\left[-\tfrac{1}{2}R\left\{\sum_{i=1}^{n} R_{ii}\frac{x_i^2}{\mu_{ii}} + 2\sum_{i}\sum_{j>i} R_{ij}\frac{x_ix_j}{\sqrt{\mu_{ii}\mu_{jj}}}\right\}\right]
$$

where

$$
R = \begin{vmatrix} 1 & r_{12} & \cdots & r_{1n} \\ r_{21} & 1 & \cdots & r_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ r_{n1} & r_{2n} & \cdots & 1 \end{vmatrix}.
$$

R_{ij} are the algebraical supplements in R, $r_{ij} = \dfrac{\mu_{ij}}{\sqrt{\mu_{ii}\mu_{jj}}}$, and χ^2 is a positive definite quadratic form in the variables $x_1, x_2, \cdots, x_n$.

Again let $u_1, u_2, \cdots, u_m$ be a system of linear forms in the variables $x_1, x_2, \cdots, x_n$

$$
\begin{aligned}
u_1 &= a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\
u_2 &= a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \qquad (11)\\
&\cdots\cdots\cdots\cdots\cdots\cdots \\
u_m &= a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n.
\end{aligned}
$$

Then from Theorem 5 follows the

COROLLARY. *The random variables* $u_1, u_2, \cdots, u_m$ *are subject to the m-dimensional properly normal distribution law of Gauss if and only if the matrix*

$$\begin{Vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{Vmatrix}$$

of the system of forms (11) *has the rank m.*

REFERENCES

[1] R. Frisch, "Correlation and Scatter in Statistical Variables," *Nordsk Statistisk Tidskrift*, vol. 1(1929), 36-102.
[2] H. Cramér, "Über eine Eigenschaft der normalen Verteilungsfunktion," *Math. Zeitschr.*, vol. 41(1936),
[3] H. Cramér, *Random Variables and Probability Distributions*, Cambridge 1937.

Moscow, U. S. S. R.

ON A CLASS OF DISTRIBUTIONS THAT APPROACH THE NORMAL DISTRIBUTION FUNCTION[1]

By George B. Dantzig

1. **Formulation of the Problem.** An important property of a sequence of binomial coefficients is that, when suitably normalized and transformed, it converges to the normal distribution.[2] The object of this paper is to exhibit a large class of other sequences which also possess this property.

The Pascal recurrence formula may be taken as the defining property of the binomial coefficients. Let the combination of n things taken x at a time be denoted by $\binom{n}{x}$. If we set $f_n(x) = (\frac{1}{2})^n \cdot \binom{n}{x}$ for $0 \leq x \leq n$ and $f_n(x) = 0$ for $x < 0$ or $x > n$, then $f_n(x)$ is defined for all integers x. With this notation Pascal's recurrence formula, $\binom{n}{x} = \binom{n-1}{x} + \binom{n-1}{x-1}$, may be written

$$f_n(x) = \tfrac{1}{2}\,[f_{n-1}(x) + f_{n-1}(x-1)], \tag{1}$$

where this new form is valid for all integers x extending from $-\infty$ to $+\infty$.

In order to generalize, we may consider a sequence of distributions $f_1(x)$, $f_2(x), \cdots, f_n(x), \cdots$ each defined in terms of the preceding one by means of the recurrence formula

$$f_n(x) = \frac{1}{a_n + 1}\,[f_{n-1}(x-0) + f_{n-1}(x-1) + f_{n-1}(x-2) + \cdots + f_{n-1}(x-a_n)], \tag{2}$$

where the x are integers, and a_n is a positive integer which may change in value from one distribution to the next. The problem is to find conditions under which $f_n(x)$, in normalized form, approaches the normal distribution. The normalization of $f_n(x)$ is effected by the affine transformation

$$u = \frac{x - \bar{x}_n}{\sigma_n}; \qquad \varphi_n(u) = f_n(x), \tag{3}$$

---

[1] Presented November 21, 1938 before a joint meeting of the Columbia Mathematics Club and the Statistical Seminar of the Graduate School of the Department of Agriculture; also December 10, 1938 before a meeting of the American Mathematical Association at the University of Maryland.

[2] Due to DeMoivre, 1731. By a variable distribution approaching the normal distribution, we mean that the integral under the variable distribution between any two limits approaches the corresponding integral under the normal curve.

where $\bar{x}_n$ and σ_n are the mean and standard deviation of the distribution $f_n(x)$. The normal (cumulative) distribution function is taken in the standard form

$$\varphi(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u} e^{-\frac{1}{2}x^2}\, dx. \tag{4}$$

The theorem whose proof forms the theme of this paper may be stated as follows:

Theorem: *A necessary and sufficient condition that $\varphi_n(u) \to \varphi(u)$ as $n \to \infty$ is that $\Gamma = 0$, where*

$$\Gamma = \lim_{n\to\infty} \sum_{i=2}^{n} \gamma_i^2 \Big/ \left(\sum_{i=2}^{n} \gamma_i\right)^2; \qquad 4\gamma_i = a_i^2 + 2a_i. \tag{5}$$

2. **Liapounoff Condition; the general case.** The recurrence formula (2) is a special case of the most general linear recurrence formula

$$f_n(x) = \sum_{i=-\infty}^{+\infty} g_n(i) f_{n-1}(x - i), \tag{6}$$

where $g_n(i)$ are a given set of weight functions generating the sequence $f_1(x)$, $f_2(x), \cdots, f_n(x), \cdots$. We may form the recurrence formula (2) by setting

$$\begin{aligned} g_n(i) &= \frac{1}{a_n + 1} && \text{if} \quad 0 \leq i \leq a_n, \\ g_n(i) &= 0 && \text{if} \quad i < 0 \text{ or } i > a_n. \end{aligned} \tag{7}$$

Let $F_k(t) = \sum_{x<t} f_k(x)$ express[3] the probability that a variable $x_k < t$, where the distribution function of x_k is defined as $f_k(x)$; and in a similar manner let the probability that a variable $s_k < t$ be given by $G_k(t) = \sum_{x<t} g_k(x)$. By summing $f_n(x)$ for all x less than t, we obtain

$$F_n(t) = \sum_{i=-\infty}^{+\infty} F_{n-1}(t - i) g_n(i) = \int_{-\infty}^{+\infty} F_{n-1}(t - i)\, dG_n(i), \tag{8}$$

where we have replaced the summation by a Stieltjes Integral. In the latter form the integral gives, in general, the probability that the *sum* of two independent variables x_{n-1} and s_n is less than t. From the above equation we see that the probability that $x_{n-1} + s_n < t$ is the same as that of $x_n < t$, so that we may set $x_n = x_{n-1} + s_n$. By iteration one obtains

$$x_n = s_1 + s_2 + \cdots + s_n \tag{9}$$

for all n. Thus we have established that *if a distribution function of a variable s_k is defined as $g_k(x)$, then the distribution function of the sum $s_1 + s_2 + \cdots + s_n = x_n$ is $f_n(x)$.*

[3] The summation extends over all values x less than t.

The limit of the distribution function of the sum of n independent variables as $n \to \infty$ has been considered by Laplace, Liapounoff, Lindeberg, and others. We shall make use of a sufficient condition given by Liapounoff that the normalized distribution function of x_n approaches $\varphi(u)$.

LAPLACE-LIAPOUNOFF THEOREM:[4] *A sufficient condition for the normalized distribution function of the sum of n independent variables $s_1, s_2, \cdots, s_n$ to approach the normal distribution function with increasing n is $\Gamma' = 0$, where*

$$\Gamma' = \operatorname*{Lim}_{n\to\infty} \frac{M_4(1) + M_4(2) + \cdots + M_4(n)}{[M_2(1) + M_2(2) + \cdots + M_2(n)]^2}, \tag{10}$$

and where $M_2(k)$ and $M_4(k)$ are defined as the second and fourth moments of s_k whose distribution is $g_k(x)$.

Thus we have shown that if a sequence of distributions $f_n(x)$ is defined by the general linear recurrence formula (6),

$$f_n(x) = \sum_{i=-\infty}^{+\infty} g_n(i)\cdot f_{n-1}(x-i),$$

then a sufficient condition that $\varphi_n(u) \to \varphi(u)$ as $n \to \infty$ is given by $\Gamma' = 0$, where $\varphi_n(u)$ is the normalized form of $f_n(u)$.

3. **Sufficiency of the Condition** $\Gamma = 0$. We may simplify the condition $\Gamma' = 0$ for the more restricted case of a sequence of distributions defined by the recurrence formula (2). In general, the second and fourth moments of $g_n(x)$ are given by

$$\begin{aligned} M_2(k) &= \sum_{x=-\infty}^{+\infty} g_k(x)(x - \bar{s}_k)^2, \\ M_4(k) &= \sum_{x=-\infty}^{+\infty} g_k(x)(x - \bar{s}_k)^4, \end{aligned} \tag{11}$$

where $\bar{s}_k$ is the mean value of the distribution. Equations (7) give the special values of $f_k(x)$; substituting these values in (11), and remembering the Bernoulli summation by which $1^p + 2^p + 3^p + \cdots + n^p$ may be expressed as a polynomial in n of degree $p + 1$, we obtain

$$\begin{aligned} M_2(k) &= \sum_{x=0}^{a_k} \frac{1}{a_k+1}\left(x - \frac{1}{2}a_k\right)^2 = \frac{1}{3}\left[\frac{a_k^2 + 2a_k}{4}\right] = \frac{1}{3}\gamma_k, \\ M_4(k) &= \sum_{x=0}^{a_k} \frac{1}{a_k+1}\left(x - \frac{1}{2}a_k\right)^4 \\ &= \frac{1}{5}\left[\frac{a_k^2 + 2a_k}{4}\right]^2 - \frac{1}{15}\left[\frac{a_k^2 + 2a_k}{4}\right] = \frac{1}{5}\gamma_k^2 - \frac{1}{15}\gamma_k; \end{aligned} \tag{12}$$

[4] J. V. Uspensky, *Introduction to Mathematical Probability* (McGraw-Hill, 1937), pages 284–292; the theorem is proved there by the method of characteristic functions.

whence by substitution in (10), Γ' becomes

$$\Gamma' = \lim_{n\to\infty} \frac{\frac{1}{5}\sum_{i=2}^{n}\gamma_i^2 - \frac{1}{15}\sum_{i=2}^{n}\gamma_i + M_4(1)}{\left[\frac{1}{3}\sum_{i=2}^{n}\gamma_i + M_2(1)\right]^2}. \tag{13}$$

Since $a_i \geq 1$, $\gamma_i \geq 3/4$, and thus $\sum_{i=2}^{n}\gamma_i \to \infty$ as $n \to \infty$, we may reduce Γ' in the limit to

$$\Gamma' = \frac{3}{5}\lim_{n\to\infty}\sum_{i=2}^{n}\gamma_i^2 \Big/ \left[\sum_{i=2}^{n}\gamma_i\right]^2. \tag{14}$$

Since $\Gamma' = \frac{3}{5}\Gamma$, the Liapounoff condition $\Gamma' = 0$ for normality becomes by (5), $\Gamma = 0$.

4. **Necessity of the Condition** $\Gamma = 0$. A necessary condition for normality can be found by noting that if $\varphi_n(u)$ approaches $\varphi(u)$, then the moments of $\varphi_n(u)$ must approach the corresponding moments of $\varphi(u)$.[5] Letting $\mu_4(n)$ be the 4th moment of $\varphi_n(u)$ and μ_4 the corresponding moment of the normal curve, a necessary condition is that $\mu_4(n) \to \mu_4$ as $n \to \infty$, and $\mu_4 = 3$. The 4th moment of $\varphi_n(u)$ may be expressed simply in terms of the moment of $f_n(x)$. If the symbol E stands for expected value, the second and fourth moments of $f_n(x)$ are $E(x_n - \bar{x}_n)^2$ and $E(x_n - \bar{x}_n)^4$ respectively, and the relationship is then

$$\mu_4(n) = \frac{E(x_n - \bar{x}_n)^4}{[E(x_n - \bar{x}_n)^2]^2} = \frac{E\left[\sum_{i=1}^{n}(s_n - \bar{s}_n)\right]^4}{\left\{E\left[\sum_{i=1}^{n}(s_n - \bar{s}_n)\right]^2\right\}^2}. \tag{15}$$

Expanding the sums by the multinomial theorem and taking the expected value of each term we obtain

$$E(x_n - \bar{x}_n)^2 = \sum_{i=1}^{n} E(s_i - \bar{s}_i)^2 + 2\sum_{i<j=1}^{n} E(s_i - \bar{s}_i)E(s_j - \bar{s}_j) = \sum_{i=1}^{n} M_2(i), \tag{16}$$

where $M_2(i)$ is the second moment of $g_i(x)$. In a similar manner we have

$$\begin{aligned} E(x_n - \bar{x}_n)^4 &= \sum_{i=1}^{n} M_4(i) + 6\sum_{i<j=1}^{n} M_2(i)M_2(j) \\ &= \sum_{i=1}^{n} M_4(i) + 3\left[\sum_{i=1}^{n} M_2(i)\right]^2 - 3\sum_{i=1}^{n} M_2^2(i); \end{aligned} \tag{17}$$

[5] Uspensky, loc. cit., pages 383–388.

whence

$$\mu_4(n) = 3 + \frac{\sum_{i=1}^{n} M_4(i) - 3\sum_{i=1}^{n} M_2^2(i)}{\left[\sum_{i=1}^{n} M_2(i)\right]^2}. \tag{18}$$

Since a necessary condition for normality is that Lim $\mu_4(n) \to \mu_4 = 3$, the fraction in the above equation must in the limit approach zero. Substituting $M_2(i) = \frac{1}{3}\gamma_i$ and $M_4(i) = \frac{1}{5}\gamma_i^2 - \frac{1}{15}\gamma_i$, we find that this ratio reduces immediately in the limit to the condition $\Gamma = 0$.

5. **Application to the Distribution of Inversions.** A frequency table may be set up for the number of permutations of n objects that give rise to a fixed number of inversions. Three objects marked 1, 2, 3 may be permuted in 6 ways:

(123), (132), (213), (231), (312), (321).

If (123) is taken as standard position, the number of inversions associated with the above set to bring each one into standard position are respectively 0, 1, 1, 2, 2, 3. Thus we pass from (321) to (123) by the following three inversions or adjacent interchanges: (312), (132), (123). Among the six permutations there is one giving rise to 0 inversions, two having 1 inversion, two having 2 inversions, and one having 3 inversions.

The distribution of inversions finds its application in a test of significance. The standard position is taken as a *hypothesis* of rank order, and the difference between an observed set of ranks and the hypothetical one is measured by the number of inversions. The distribution may then be used for finding the probability of obtaining by chance the number of inversions found, or less. For a moderate number of ranks (six or more), the distribution of inversions may be approximated by a normal curve. We shall show that as the number of ranks is increased, the normalized distribution of inversions approaches the normal distribution. The distribution of inversions of 1, 2, 3, 4, objects will be found in the table below.

Inversions: x	0	1	2	3	4	5	6
$1\cdot f_1(x)$	1						
$1\cdot 2\cdot f_2(x)$	1	1					
$1\cdot 2\cdot 3\cdot f_3(x)$	1	2	2	1			
$1\cdot 2\cdot 3\cdot 4\cdot f_4(x)$	1	3	5	6	5	3	1

By induction one may show that the following relationships hold between successive distributions:

$$(19)\quad \begin{aligned} f_2(x) &= \tfrac{1}{2}[f_1(x-0)+f_1(x-1)], \\ f_3(x) &= \tfrac{1}{3}[f_2(x-0)+f_2(x-1)+f_2(x-2)], \\ &\vdots \\ f_n(x) &= \frac{1}{n}[f_{n-1}(x-0)+f_{n-1}(x-1) \\ &\qquad + f_{n-2}(x-2)+\cdots+f_{n-2}(x-n+1)]. \end{aligned}$$

Since this satisfies the basic recurrence formula (2), where $a_n = n-1$, we may find out whether the normalized distributions of inversions approaches $\varphi(u)$. With $\gamma_n = n^2 - 1$ the condition $\Gamma = 0$ becomes $\lim_{n\to\infty} \sum_{i=2}^{n} (i^2-1)^2 \Big/ \left[\sum_{i=2}^{n}(i^2-1)\right]^2$. The numerator sums to a polynomial of the 5th degree in n, while the brackets of the denominator sums to a 3d degree polynomial, which after squaring is of the 6th degree; so that as $n \to \infty$ we have in the limit $\Gamma = 0$. Thus the normalized distribution function of the inversions of n objects approaches $\varphi(u)$ as $n \to \infty$.

Equations (12) and (16) permit us to find the mean and standard deviation of the distribution of the inversions of n objects:

$$(20)\quad \begin{aligned} \bar{x}_n &= \tfrac{1}{4}n(n-1), \\ \sigma_n^2 &= \frac{1}{72}n(n-1)(2n+5). \end{aligned}$$

The sequence of binomial coefficients, and the distribution of inversions are examples of sequences that satisfy recurrence relation (2); it should be noted that their respective values of γ_n, ($\gamma_n = 3/4$ or $\gamma_n = n^2 - 1$), may be considered as *bounded* between two polynomials of the same degree in n. Whenever this is true the condition $\Gamma = 0$ will hold and $\varphi_n(u)$ will approach $\varphi(u)$. On the other hand, if for example, $\gamma_n = r^n$, then $\Gamma \neq 0$ and $\varphi_n(u)$ does not approach $\varphi(u)$.

6. **Smoothing Formulas.** The general recurrence formula (6),

$$f_n(x) \doteq \sum_{i=-\infty}^{+\infty} g_n(i) f_{n-1}(x-i),$$

may be considered as a linear smoothing formula. For example, we may obtain the usual three point smoothing formula based on binomial coefficients for smoothing a distribution $f_1(x)$ into $f_2(x)$ by setting in the above equation $n = 2$, $g_2(i) = \frac{1}{4}\binom{2}{i+1}$ for $-1 \le i \le +1$, and $g_2(i) = 0$ for $i < -1$ or $i > +1$. Thus

$$(21)\quad f_2(x) = \tfrac{1}{4}[f_1(x+1) + 2f_1(x) + f_1(x-1)].$$

From considerations found in Section 2, we see that if a variable x_1 has for distribution $f_1(x)$ and a variable s_2 has for distribution $g_2(x)$, then their *sum* $s_2 + x_1$ has for distribution function the smoothed distribution $f_2(x)$. From this point of view, the smoothed distribution $f_2(x)$, obtained by applying a linear smoothing formula, is a "cross" between the original unsmoothed distribution $f_1(x)$ and the artificial weight distribution $g_2(x)$.

Often a smoothing formula is used several times; first on the original distribution, then on the smoothed distribution, and then sometimes on the smoothed smoothed distribution. *If a linear smoothing formula is thus iterated* 1, 2, 3, $\cdots$, n, $\cdots$ *times, the sequence of smoothed distributions obtained upon normalization approaches* $\varphi(u)$. This may easily be demonstrated by showing that Liapounoff's condition for normality, $\Gamma' = 0$, is satisfied. Since in this case the weight distribution $g_n(i)$ is the same for all $n \geq 2$, the corresponding moments of these distributions must all be equal; thus we may write $M_4(n) = M_4(2)$ and $M_2(n) = M_2(2)$ where $n \geq 2$. Substituting in (10), we obtain for Γ'

$$\Gamma' = \lim_{n\to\infty} \frac{M_4(1) + (n-1)M_4(2)}{[M_2(1) + (n-1)M_2(2)]^2}, \tag{22}$$

where $M_2(1)$ and $M_4(1)$ are the 2d and 4th moments of the unsmoothed distribution $f_1(x)$. The mean value $\bar{x}_n$ and the standard deviation σ_n of the distribution $f_n(x)$ formed by iterating a smoothing formula $n - 1$ times are easily shown to be

$$\begin{aligned} \bar{x}_n &= \bar{x}_1 + (n-1)\bar{s}_w, \\ \sigma_n^2 &= \sigma_1^2 + (n-1)\varphi_w^2, \end{aligned} \tag{23}$$

where $\bar{x}_1$ and σ_1 are the mean and standard deviations of the original unsmoothed distribution, and where $\bar{s}_w$ and σ_w are the mean and standard deviation of the weight distribution $g_2(i)$.

The linear smoothing formula is used in practical work to smooth data. Successive application of one or many such linear formulas will usually smooth *any* set of values to the normal curve of error. The above section serves as a warning of what is introduced by the use of such methods.

It is a pleasure to acknowledge the helpful criticisms and advice of Dr. W. E. Deming in the preparation of the manuscript.

WASHINGTON, D. C.

THE LENGTH OF THE CYCLES WHICH RESULT FROM THE GRADUATION OF CHANCE ELEMENTS

By Edward L. Dodd

1. **Introduction.** Eugen Slutzky[1] found that under certain conditions repeated summations of chance elements lead to a sinusoidal configuration. Generalizations were made by V. Romanovsky.[2] A more recent paper by Slutzky[3] has appeared, summarizing his original Russian memoir, and making extensions. Contributions to this subject have also been made by H. E. Jones,[4] E. J. Moulton,[5] and A. Wald.[6]

Readers who wish to get into touch with recent literature on periodicity are referred to two excellent books, that of Karl Stumpff[7] with bibliography of 319 references, and that of Herman Wold,[8] with bibliography of nearly 70 references.

In this paper, I deal with the wavy configuration resulting from a *single* application of a specified graduation formula. For this purpose, only linear operators are considered. For actual graduation it is customary to require that the sum of the coefficients or "weights" be equal to unity. But for the present purpose, this requirement is irrelevant. For example, summing and averaging are here essentially identical. The graduation formula considered may or may not be the combination of simple summations or averages. Indeed, formulas preferred by actuaries and statisticians include terms with *negative* coefficients; and thus involve an operation other than addition. F. R. Mac-

---

[1] Eugen Slutzky, "Sur un théorème limite relatif aux series des quantités éventuelles." *Comptes Rendus*, Vol. 185 (1927) pp. 169–171.

[2] V. Romanovsky, "Généralisations d'un théorème de M. E. Slutzky." *Comptes Rendus*, Vol. 192(1931) pp. 718–721. "Sur la loi sinusoidale limite." *Rendiconto Circolo Mathematico di Palermo*, Vol. 56 (1932) pp. 82–111. "Sur une généralisation de la loi sinusoidale limite." Ibid., Vol. 57 (1933) pp. 130–136.

[3] E. Slutzky, "The summation of random causes as a source of cyclic processes." *Econometrica*, Vol. 5 (1937) pp. 105–146.

[4] H. E. Jones, "The theory of runs applied to time series," *Report of Third Annual Conference of Cowles Commission for Research in Economics* (1937) pp. 33–36. This abstract itself does not include reference to repetitions, mentioned by Moulton and Wald.

[5] E. J. Moulton, "The periodic function obtained by repeated accumulation of a statistical series. *American Mathematical Monthly*, Vol. 45 (1938), pp. 583–586.

[6] A. Wald, "Long cycles as a result of repeated integration." *American Mathematical Monthly*, Vol. 46 (1939), pp. 136–141.

[7] Karl Stumpff, *Grundlagen und Methoden der Periodenforschung*, Berlin, 1937, Julius Springer.

[8] Herman Wold, *A Study in the Analysis of Stationary Time Series*. Uppsala, 1938, Almqvist and Wiksells.

aulay[9] gives a chart of 24 weight diagrams. Of these only the first four are without negative coefficients.

Of course, the "waves" produced are irregular, and the difficulty of defining a cycle-length confronts us. The apparently naïve definition of a cycle-length as the average distance between successive maxima (or minima) is, I believe, worth consideration as a rough first approximation of the cycle length for graduated values delivered by formulas with negative coefficients or by those involving at least triple summations. But the cycle length thus determined is somewhat too short; for, slight undulations will occur—Slutzky calls them "ripples"—which should be eliminated if we want a cycle-length *intuitionally reasonable*. On the other hand, the cycle-length defined as the average distance between alternate intersections of the graduated curve with the base line is likely to be decidedly too long,—as illustrated by Slutzky's Figure 2 (loc. cit., p. 109) which exhibits 1,000 graduated items, with 41 marked maxima and 41 marked minima—after *elimination* of what he considers ripples—but with only 23 up-crossings and 23 down-crossings of the base line. I indicate in what follows an analytic method for removing ripples. And I describe *several methods* for obtaining a number which might be called a cycle-length. Often these seem to *cluster about a central value*, which appears to me to be a reasonable estimate of the "*length of the cycle*" created by the *specified graduation formula*.

The *theory* to be presented here assumes that the chance elements are *normally distributed* about zero with constant variance. But the data used by Slutzky came from lottery drawings, with a "rectangular" distribution; and for checking I have likewise used *rectangular distributions*; mainly, three sets of 600 numbers each, taken from the tenth figures of logarithms in the Vega Tables. It is known, however, that the average of a few elements distributed rectangularly is nearly normal. From many tests that I have made, it would seem that rectangular distributions react as if normal. To illustrate: When normal data are given a twelve-fold summation or averaging by twos, the probabilities that at a specified point there will be an upcrossing of the base line, a maximum, or an inflection from concave to convex are respectively, 0.0628, 0.106, and 0.134. These numbers multiplied by 100 give 6.28, 10.6, and 13.4, as the expected number of occurrences per hundred graduated values. Slutzky exhibits in Figure 4 (loc. cit., p. 111) 100 ordinates as the result of applying to lottery drawings 12-fold summation by twos. The figure shows 6 or 7 up-crossings, ten maxima, and 13 or 14 such inflections—in close agreement with *expectations based upon normal distributions.*

2. Derivation of Probabilities and Comparison of Actual with Expected Occurrences. A "cycle length" is first conceived of as the *reciprocal of a relative frequency or probability*. Thus, if the probability that a graduated value will

[9] F. R. Macaulay, *The Smoothing of Time Series*. Publications of the National Bureau of Economic Research, incorporated, No. 19 (1931). See pp. 77–79.

be a maximum is 0.05, we expect 5 maxima per hundred graduated values, making the "cycle length" for maxima equal to 20. It will be recalled that if p is the probability of an occurrence of an event in a single trial, then in s trials the expected number of occurrences is sp, whether the trials are *independent or not.*

It is assumed that the *data,* $x_1, x_2, \cdots$ are *independent and normally distributed about zero with constant variance* V. Then any linear function

$$(1) \qquad y_r = a_{-m}x_{r-m} + \cdots + a_0 x_r + a_1 x_{r+1} + \cdots + a_m x_{r+m}$$

is likewise normally distributed about zero; and the variance of y_r is $V = \Sigma a_i^2$.

(a) *Probabilities When Two Conditions Are Imposed.* Consider first the "planes" $y_{r-1} = 0$ and $y_r = 0$, each in $2m + 1$ dimensions; and jointly in $2m + 2$ dimensions. They form four "dihedral" angles. Let

$$(2) \qquad \theta = \text{angle between } y_{r-1} = 0 \quad \text{and} \quad y_r = 0,$$

the inside points $(x_{r-m-1}, \cdots, x_{r+m})$ being such that $y_{r-1} < 0$, and $y_r > 0$. Now, an orthogonal transformation or "rotation "leaves invariant this angle θ and also the normal probability function:

$$(3) \qquad \text{Probability} = \text{Const.} \cdot \exp\left[-\Sigma x_i^2/2V\right].$$

The angle θ may be found[10] from

$$(4) \qquad \cos\theta = \sum_{i=-m}^{m-1} a_i a_{i+1} \Big/ \sum_{-m}^{m} a_i^2.$$

Let us think of the rotation which carries the intersection of the planes into the "vertical" position. To find the probability that $y_{r-1} < 0$ and $y_r > 0$, we integrate over all $2m + 2$ dimensional space which lies between the two planes in the dihedral angle thus characterized. For $2m$ of such variables, the integration extends from $-\infty$ to $+\infty$ yielding unity as a factor. If u and v are the two variables that remain, then we are to find the volume of that portion of the solid

$$(5) \qquad z = (1/2\pi V) \exp\left[-(u^2 + v^2)/2V\right]$$

which lies between two vertical planes through the origin making the angle θ with each other. Then,

$$(6) \qquad \text{Probability of up-crossing} = \theta/360°.$$

$$(7) \qquad \text{Cycle length for up-crossing} = 360°/\theta.$$

Let

$$\Delta y_r = y_{r+1} - y_r.$$

[10] D. M. Y. Sommerville. *An introduction to the Geometry of N Dimensions.* Methuen and Co., Ltd., London, 1929. See p. 76.

Then y_r is a maximum if $\Delta y_{r-1} > 0$ and $\Delta y_r < 0$. Suppose

$$\theta_1 = \text{angle between } \Delta y_{r-1} = 0 \quad \text{and} \quad \Delta y_r = 0, \tag{8}$$

inside points making $\Delta y_{r-1} > 0$ and $\Delta y_r < 0$. Then

$$\text{Probability for maximum at } y_r = \theta_1/360° \tag{9}$$

$$\text{Cycle length for maxima} = 360°/\theta. \tag{10}$$

The same equations apply to minima; since for minima we simply reverse the two foregoing inequalities, and pass to the equal "vertical" dihedral angle.

Likewise, from $\Delta^2 y_{r-1} < 0$ and $\Delta^2 y_r > 0$ we obtain an angle θ_2 such that $\theta_2/360°$ is the probability for change of inflection from concave downward to convex downward. This is also equal to the probability for change of inflection from convex to concave. Such changes of inflection have some interest on their own account and in checking; but do not seem to have any direct bearing upon the main problem under discussion here.

(b) *Probabilities When Three Conditions Are Imposed.* We consider now the *elimination of ripples.* To make y_r a maximum, two linear conditions are required. A third linear condition such as $y_r > \frac{1}{2}(y_{r-k} + y_{r+k})$, or simply $y_r > y_{r+k}$, with $k > 1$, will remove some ripples. Suppose we have given three planes through the origin,

$$\begin{aligned} a_1x_1 + a_2x_2 + \cdots + a_nx_n &= 0, \\ b_1x_1 + b_2x_2 + \cdots + b_nx_n &= 0, \\ c_1x_1 + c_2x_2 + \cdots + c_nx_n &= 0. \end{aligned} \tag{11}$$

The angles between these planes in pairs are

$$\cos\alpha = \frac{\Sigma b_i c_i}{(\Sigma b_i^2 \cdot \Sigma c_i^2)^{\frac{1}{2}}}; \qquad \cos\beta = \frac{\Sigma a_i c_i}{(\Sigma a_i^2 \cdot \Sigma c_i^2)^{\frac{1}{2}}}; \qquad \cos\gamma = \frac{\Sigma a_i b_i}{(\Sigma a_i^2 \cdot \Sigma b_i^2)^{\frac{1}{2}}} \tag{12}$$

In general, eight-trihedral angles are thus formed at the origin; since we may take acute angles for α, β, and γ or their supplements. By an orthogonal transformation or "rotation about the origin" we are led to the three dimensional problem of finding the portion of a sphere lying in a specified spherical pyramid with base a spherical triangle, ABC, having spherical excess $E = A + B + C - 180°$. Now the spherical surface is 4 great circles or 720°. Hence, for a maximum, subject to an additional linear homogeneous inequality,

$$\text{Probability of conditioned maximum} = E/720° \tag{13}$$

care having been taken to enter the proper trihedral angle.

(c) *Probabilities When Four Conditions Are Imposed.* To avoid complexities involved in the use of four intersecting planes, the following expedient was employed. Consider a set of values of y_r such that this y_r is a maximum. Among these there is theoretically a certain fraction or proportion p at which also

$y_r > y_{r+k}$, with $k > i$, and the same proportion p satisfying $y_r > y_{r-k}$. Let p' be the proportion satisfying both inequalities. Then $1 - p' \leqq 1 - p + 1 - p$ leads to

$$p' \geqq 2p - 1 = p^2 - (1 - p)^2. \tag{14}$$

If p is fairly close to unity; a good approximation for p' would seem to be

$$p' = p^2. \tag{15}$$

This p^2 would have been exact for p', had the graduated values been independent. That p' is here only slightly above $2p - 1$ seems likely, from the graduations that I have examined; for, the failure of one of the inequalities $y_r > y_{r+k}$ or $y_r > y_{r-k}$ was seldom accompanied by the failure of the other.

For graduations with the Spencer 21-term formula, when $k = 5$, $p = 0.936$, and $(1 - p)^2 = 0.0041$, which is fairly small. In practice, we would find in this case directly $P = 0.07125$ = probability of a maximum; $Pp = 0.0668$ = probability of a maximum at y_r with $y_r > y_{r+5}$. Then the probability Pp' of a maximum at y_r with $y_r > y_{r+5}$ and $y_r > y_{r-5}$ would have as lower bound $2Pp - P = 2(0.0668) - 0.07125 = 0.06235$.

But a closer approximation to the actual value would seem to be $Pp^2 = (Pp)^2/P = (0.0668)^2/0.07125 = 0.0626$.

This would give a cycle length of $1/0.0626 = 15.97$.

(d) *Indications from Correlation Theory.* We may also attempt to estimate a cycle length with the aid of correlation theory. If for graduation, we use a linear operator with coefficients proportional to successive ordinates of a cosine curve with a specified period, it is, I presume, fairly well known that the graduated values tend to exhibit the period of that cosine curve. Moreover, this quasi period may be induced very strongly with the use of formulas which represent "damped vibration" as shown by H. Labrouste[11] and Mrs. Labrouste. Now many standard graduation formulas have plots resembling somewhat damped vibration. Here, the central symmetrical arch leading down to the lowest negative terms on each side is usually large in comparison with the flanking waves. Now for a *strict cosine* curve of period $2k$, the coefficient of correlation of y_r and y_{r+k} is -1, at least theoretically. For *chance* material y_r, with mean zero and constant variance, the coefficient of correlation between y_r and y_{r+j} is defined in terms of expected values, thus:

$$\rho_j = E(y_r y_{r+j})/E(y_r^2). \tag{16}$$

For graduated values, y_r, we might then *seek the value j which will make ρ_j as close to -1 as possible.* But for most common graduation formulas, ρ_j does not approximate -1. This difficulty, however, disappears if the graduation

[11] H. and Mrs. Labrouste, "Harmonic analysis by means of linear combinations of ordinates," *Terrestrial Magnetism and Atmospheric Electricity*, Vol. 41 (1936) pp. 15–28. See pp. 17, 18.

formula is properly centered. In a Fourier series, there is a constant term, which is of no significance in indicating oscillations, and is sometimes eliminated. The analogous modification for a linear graduation formula with n coefficients—of which the sum is unity—would seem to be the subtraction of $1/n$ from each coefficient, forming what I regard as a *residual. For this residual, negative correlations of substantial size appear.* And that j with which the numerically largest negative correlation ρ_j is associated may be considered as indicating a half-cycle length.

In the case of the Spencer 21-term formula, $j = 8$, making cycle-length $= 16$, just about identical with the cycle length for maxima at y_r with $y_r > y_{r-5}$ and $y_r > y_{r+5}$.

(e) *The period of a Closely Fitting Cosine Curve.* By another route, also, we may approach the problem of associating with a specified linear graduation a number as the length of induced cycles. We shall consider here only those formulas in which the coefficients are symmetrical with respect to the center. In equation (1), this means that $a_{-j} = a_j; j = 1, 2, \cdots, m$. Suppose now that the x's are no longer chance elements, but are the successive terms of a cosine curve with period k. That is:

$$x_r = \cos(r\theta + \alpha); \qquad \theta = 2\pi/k = 360°/k. \tag{17}$$

Then, if $a_{-j} = a_j$, it follows that

$$a_{-j}x_{r-j} + a_jx_{r+j} = 2a_j \cos j\theta \cdot \cos(r\theta + \alpha). \tag{18}$$

Then, from (1),

$$y_r = C \cos(r\theta + \alpha), \tag{19}$$

where C is independent of r. For a given graduation formula, with a's specified, this C depends upon θ, or we may say, upon $k = 360°/\theta$. *We may regard the graduation formula* y_0 *as "fitting best" the curve* $\cos[r(360°)/k]$ *when* k *is so chosen as to give to* C *a largest value.* The presumption is that the graduation formula will curl chance data up into cycles in about the same fashion as a cosine curve to which it is closely akin. The actual period of this closely fitting cosine curve may then be taken as the quasi-period or "cycle-length" of the graduation formula.

If, relying upon intuition, we were to select a cosine curve to fit a given graduation formula, we might easily decide to disregard the small waves that usually flank the central arch, and to take a cosine curve with a span—distance between minima—equal to the span of this central arch. *In fact, this span gives, I believe, a good first estimate of the cycle length of the induced waves. This first estimate seems, however, a trifle too small.*

3. Size of Ripples, Simple Summation, Variability, and Height of Waves.

(a) *Size of Ripples.* In the use of $y_r > y_{r+k}$ to remove ripples, what integer should we take for k? The dividing line between ripples and waves is of course

arbitrary. As Figure 2, p. 109, Slutzky exhibits 1,000 graduated values from two-fold summations by 10, with ripples removed. He states (p. 119): "maxima and minima with amplitudes of ten units or less being discarded as ripples." For this double summation, I find that the probability that y_r will be a maximum with $y_r > y_{r+10}$ and $y_r > y_{r-10}$ is approximately 0.0437. Among 1,000 graduated values, 43.7 such maxima would then be expected. Slutzky marks with arrows the 41 maxima which remain after the elimination of what he regards as ripples. The reciprocal of 0.0437 gives 22.9 as cycle length. Then $k = 10$ is less than half this cycle-length. For standard graduation formulas, it would seem likely that a *value of k about one-third the span of its central arch would eliminate fairly well the inconsequential fluctuations*; and likewise for graduations, with coefficients forming an arch with nearly horizontal ends, like twelve-fold summation by twos, with arch span 12. For this twelve-fold summation, I find that 0.0831 is the probability that a maximum will occur at y_r, with $y_r > y_{r+4}$ and $y_r > y_{r-4}$, giving 8.31 such maxima per hundred graduated values. Slutzky's Figure IVa shows eight such maxima, and two ripples.

(b) *Simple Summation.* I shall not discuss in detail the cycles produced by simple summation or averaging. Formulas for probability here are relatively simple. Thus, for the sum or average of n normal chance data, the probability of a maximum is 1/4, irrespective of the value of n. This appears to be about valid for rectangular data if we count the weak maxima. *A simple average of chance data, however, seems to inherit largely the chaotic character of the present data. But some sinuosity is, after all, implanted.*

(c) *Variability.* A general discussion of the variability of induced waves is beyond the scope of this paper. However, I record a numerical result. For the Spencer 21-term graduation formula, the probability of a maximum is 0.07125. Among 580 graduated values, then, 41.3 maxima would be expected. Actually, 42 maxima were found. Now, if $n - 1$ points are placed "at random" on a line of unit length—here dx is the probability that a point will fall in an interval of length dx—then the expected value[12] of the sum of the squares of the resulting n segments is $2/(n + 1)$. Thus, if 42 points are placed at random on an interval of 580 units, the expected sum of the squares of the segments is $(2/44)\ (580)^2 = 15{,}290.9$. But, if the points are placed at equal intervals, this sum of squares takes its least value, $(580)^2/43 = 7{,}823.3$. Then, $15{,}290.9 - 7{,}823.3 = 7{,}467.6$. On the other hand, the 42 maxima among *Spencer graduated values* gave segments for which the sum of the squares was 8,656.5; that is, *only* 833.2 *in excess of the above* 7,823.3, *which represents perfect periodicity for maxima.* Of course, this excess of 833.2 indicates considerable departure from perfect periodicity; but it is nowhere near the 7,467.6 to be expected from a random distribution of points. In spite of irregularities, *the sinusoidal character of graduated values is conspicuous.*

(d) *The Height or Amplitude of Induced Waves.* While our chief interest

[12] W. Burnside, *Theory of Probability*, Cambridge University Press, 1928. See p. 71.

here lies in what is called the *length* of a cycle, a brief reference may well be made to the amplitude or *height* of the induced waves. The operation of the linear function y_r in (1) upon data with variance V yields graduated values with variance $V\Sigma a_i^2$. This particular statement does not require the assumption of normality. Thus the Spencer 21-term formula is expected to produce graduated values with a standard deviation 37.8% of that of the data. This represents some reduction, of course; but, nevertheless, the *"waves" stand out in bold relief. They are not diminutive.*

4. **Data and Graduations Examined.** Slutzky's graduations, exhibited in *Econometrica*, Vol. 5, have already been mentioned. Three sets of chance data were graduated by students at the University of Texas, Mr. Victor W. Pfeiffer in 1936, Mr. C. M. Tolar and Miss Anna Velma Martin, in 1938, to make tests with regard to smoothing coefficients,[13] the results appearing in M.A. theses. The data were figures in the tenth place of the Vega logarithm tables, 600 numbers in each set, as follows: Logarithms from 200 to 799; logarithms of cosines of angles from 0° to 59° 54′, by intervals of 6′; logarithms of sines of angles from 6′ to 60° by intervals of 6′. The graduation formulas used were all symmetric, with $a_{-j} = a_j$. Mr. Pfeiffer used the Spencer 21-term formula, with coefficients 1/350 of:

$$-1, -3, -5, -5, -2, 6, 18, 33, 47, 57, 60, 57, \text{etc.}$$

The other two formulas used were 11-term formulas which I devised, correct to third differences, and with fourth differences rather small, described by: $-1.13\, D^4$ and $-0.97\, D^4$, where $D = \log_e E$ (see Henderson, loc. cit., pp. 26–37); as compared with $-5.4\, D^4$ for Woolhouse 15-term, and $-12.6\, D^4$ for Spencer 21-term. These two 11-term formulas are:
(i) Averaging by twos, threes, and fours, applied to (1/12) (−4, 3, 14, 3, −4) yielding (1/288) (−4, −9, 3, 36, 73, 90, 73, 36, 3, −9, −4);
(ii) Triple averaging by threes, applied to (1/10) (−3, 2, 12, 2, −3) yielding (1/270) (−3, −7, 0, 29, 71, 90, 71, 29, 0, −7, −3). From part of the foregoing data, also, I made other graduations to check certain probabilities.

5. **Cycle Lengths for the Spencer 21-Term Graduation Formula.** All the various determinations of cycle length mentioned in the foregoing were applied to the Spencer formula, and to some other formulas. The results obtained for the Spencer formula seem representative, and will be given here in detail. Our main conception of a cycle-length is that it is the reciprocal of a probability or relative frequency. The probability of a minimum is the same as that of a maximum; of a down-crossing of the base line, the same as that of an up-crossing. Probabilities are listed that the representative ordinate y_r will be a maximum—

[13] Robert Henderson, *Graduation of Mortality and Other Tables*, Actuarial Society of America, New York, 1919, p. 34.

with or without further restrictions. The probability is given for an up-cross at the representative abscissa x_r. In the table which follows, a middle entry for a cycle length of 16 is obtained from the "residual" described in (d) of Section 2.

The Expected Length of Cycles Produced When Normal Chance Data Are Graduated by the Spencer 21-term Formula in Accordance with Various Specifications for the Cycle

Specification	Probability	Cycle-Length
Maximum at y_r	0.07125	14.0
Maximum at y_r with $y_r > y_{r+5}$	0.0668	15.0
Maximum at y_r with $y_r > \frac{1}{2}(y_{r-7} + y_{r+7})$	0.0657	15.2
Maximum at y_r with $y_r > y_{r+5}$, and $y_r > y_{r-5}$	0.0626	16.0
By use of "residual". (See 2(d))		16.0
Maximum at y_r with $y_r > y_{r+7}$	0.0623	16.1
Period of "best fitting" cosine curve. (See 2(e))		16.7
Maximum at y_r, $y_r > 0$. (Or: $y_r >$ Mean y_r)	0.0591	16.9
Maximum at y_r with $y_r > y_{r+7}$ and $y_r > y_{r-7}$	0.0545	18.3
Up cross at x_r	0.0469	21.3

The foregoing exhibit seems to suggest a cycle length of something like 16 *for the cycles created by the operation of the Spencer* 21-*term formula upon chance data.* This is just about the reciprocal of the probability that a maximum will occur at y_r with $y_r > y_{r+5}$ and $y_r > y_{r-5}$. If 16 is thus set up as the standard wave length, a wave of 10 units extending from x_{r-5} to x_{r+5} would not be regarded as insignificant.

Now 16 is also the interval between the outermost low coefficients, -5, in the Spencer formula. The plot of a curve through ordinates equal to the Spencer coefficients would probably make the central arch have a span of about 15. This 15 seems a little too small as a representative of cycle lengths obtained by the foregoing different methods.

From the theory set forth, 0.0626 is the probability that a maximum will occur at y_r with $y_r > y_{r+5}$ and $y_r > y_{r-5}$. Then among 580 graduated values, 36.3 such maxima would be expected. Among the Pfeiffer graduated values 38 were actually found.

6. Comparative Results of Seven Graduation Formulas. An exhibit will now be made of results obtained from seven graduation formulas. Of these, the simplest is double averaging or summation by tens, with coefficients forming a

triangular arch, with a "span" which will be set down here as 18. Next in order of simplicity—avoiding negative coefficients—is 12-fold averaging by twos. Probabilities are given that a maximum will occur at a point y_r, with $y_r > y_{r-k}$ and $y_r > y_{r+k}$ for what seems to be appropriate values of k. In the five cases where graduations were made, the number of the maxima of specified character actually found are set down in line with their expected values. Also the span of the central arch is compared with cycle lengths.

Macaulay (loc. cit., pp. 73, 74) mentions favorably a 43-term formula obtained as follows: Summation by 8, by 12, doubly by 5, applied to weights: +7, −10, 0, 0, 0, 0, 0, 0, +10, 0, 0, 0, 0, 0, 0, −10, +7. This is the longest formula to be considered here.

As noted before, my theory is based upon the assumption of a normal distribution for the data. The data actually tested had a rectangular distribution. Nevertheless, close agreement was found between the expected number of maxima and the number actually found.

Results of Applying Seven Graduation Formulas to Chance Data. Comparison of the Expected Number of Conditioned Maxima with the Actual Number Found Among Graduated Values, and Comparison of Cycle Length with Span of Central Arch

(1) Graduation Formula	(2) k	(3) Probability Max. at y_r $y_r > y_{r-k}$ $y_r > y_{r+k}$	(4) Number of Graduated Items, y_r	(5) Expected Number of Such Maxima	(6) Actual Number of Such Maxima	(7) Cycle as 1/(3)	(8) Span of Central Arch
11-term by Tolar	3	0.110	590	64.9	67	9.09	8
11-term by Martin	3	0.114	590	67.3	65	8.77	8
13-term $(2)^{12}$ by Slutzky	4	0.0831	100	8.31	8	12.0	12
19-term $(10)^2$ by Slutzky	10	0.0437	1,000	43.7	41	22.9	18
21-term Spencer by Pfeiffer	5	0.0668	580	36.3	38	16.0	15
29-term Kenchington	8	0.0428				23.4	20
43-term Macaulay	9	0.0389				25.7	22

7. **Summary.** E. Slutzky found that the summing of chance data resulted in series of numbers with something like a cyclic appearance,—this being intensified by repetition of the summing. Slutzky and others have proven limit theorems. In this paper, I study the effects of a single application of a graduation process upon chance data. The most acceptable graduation formulas contain negative coefficients, and thus involve something more than repeated summations. Several methods are discussed for assigning to a given graduation formula a number as the length of the cycles it tends to produce. One of the most satisfactory of these is in line with the suggestion of Slutzky that before counting maxima, any insignificant "ripples" should be eliminated. The proba-

bility is found that a graduated value y_r should be a maximum—greater than the two adjacent values y_{r-1} and y_{r+1}—with the further condition that for some appropriate k, y_r shall be greater than y_{r-k} or y_{r+k} or both. The reciprocal of this latter probability is suggested as the length of the cycle which the given graduation tends to implant in the graduated values.

THE UNIVERSITY OF TEXAS.

ON THE DISTRIBUTION OF THE "STUDENT" RATIO FOR SMALL SAMPLES FROM CERTAIN NON-NORMAL POPULATIONS[1]

By H. L. Rietz

Much of interest in the theory and practice of statistical methods has been developed around the distribution function,

$$\frac{\Gamma(N/2)}{\pi^{\frac{1}{2}}\Gamma\left(\frac{N-1}{2}\right)(1+z^2)^{N/2}} \tag{1}$$

of the "Student" ratio, $z = \dfrac{\bar{x} - m}{s}$, where $\bar{x}$ denotes the mean, s the standard deviation of a sample of N items, say $x_1, x_2, \cdots, x_N$, taken at random from a normally distributed parent population of mean, m.

The investigations of certain non-normal parent distributions by Shewhart and Winters [1], Rider [2], E. S. Pearson [3], M. S. Bartlett [4], and R. C. Geary [5] indicate that applications of the "Student" theory give more satisfactory results than the classical theory for a considerable variety of non-normal parent distributions, but some of these investigators find that the theory fails in certain cases to describe the facts to an extent that suggests further experimental sampling investigations along this line whenever suitable data are available. Others infer that a completely satisfactory analysis of the position of the "Student" z-test will be possible only if the theoretical distribution of z in samples from the non-normal distribution in question becomes known. Several of the above named statisticians have attributed the failures of the distribution (1) to describe their data, in large part, to the correlation between $x = \bar{x} - m$ and s. For this reason, there is considerable interest in the degree of correlation between $x = \bar{x} - m$ and s, and especially in the nature of the regression of s or of s^2 on x.

The present paper gives an analysis of data obtained by experimental sampling from two non-normal distributions whose sources we shall now describe. The parent distributions with which the paper is concerned are theoretical distributions resulting from certain urn schemata devised [6] by the writer some years ago.

In 1925, Leone E. Chesire, in an unpublished thesis for the degree, Master of Science, at the University of Iowa, obtained data by experimental sampling, that seem to be appropriate material for a study of the correlation of mean and standard deviation for small samples from certain non-normal distributions.

One of the original bivariate parent populations, whose marginal totals we are

---

[1] Presented in part before the American Mathematical Society under a somewhat different title, November 26, 1937.

using, exhibited linear regression while the other exhibited non-linear regression. For convenience in distinguishing between the two cases, we shall speak of material from the linear case as Case I and that from the non-linear case as Case II. After devising a scheme for drawing pairs of variates at random, 5,000 pairs were drawn in sets of five for each of the two cases.

While the primary purpose of this experimental sampling was to study the distributions of means, standard deviations, and correlation coefficients [7] for small samples from the non-normal populations, we have as a by-product, in the marginal totals of the correlation tables, four sets of 1,000 pairs of means and standard deviations. However, since three of the four sets of marginal totals of the two theoretical parent correlations tables are alike, we have actually only two significantly different sets to consider.

Case I. For the case of linear regression of y on x in the bivariate parent population, the parent distribution from the marginal totals may be very simply described by showing the frequency distribution in Table 1.

TABLE 1

Sums in second throw of dice-values of stochastic variable	2	3	4	5	6	7	8	9	10	11	12
Frequency	6	12	18	24	30	36	30	24	18	12	6

The moment coefficients and β's which characterize the distribution given in Table 1 are:

$$\text{Mean} = 7, \qquad \mu_2 = 5\tfrac{5}{6}, \qquad \mu_3 = 0, \qquad \mu_4 = 80.5, \qquad \beta_1 = 0, \qquad \beta_2 = 2\tfrac{64}{175}.$$

Each of the 1000 sets of five drawn from the distribution in Table 1, yields a mean $\bar{y}$ and a standard deviation, s_y, which we shall denote by w to make our notation simpler to write. Table 2 is the correlation table of the pairs $(\bar{y}, w)$. The correlation coefficient $r_{w\bar{y}}$, between mean $\bar{y}$ and standard deviation $s_y = w$ has a value

$$r_{w\bar{y}} = -0.020 \pm 0.021$$

which differs insignificantly from zero.

The uncorrected value of the correlation ratio of w on $\bar{y}$ is

$$\eta_{w\bar{y}} = 0.182.$$

When we remember that the correlation ratio is not free to vary in the negative direction from 0, and apply the Pearson correction [8] for this situation together with the "Student" correction [9] for grouping, we obtain for the corrected, $\eta_{w\bar{y}}$, the value 0.133.

It becomes fairly obvious that significant correlation exists and that the regression is non-linear. Indeed, it has been shown recently by Geary [5, pp. 178–9] that normality in the parent distribution is both a necessary and

TABLE 2

Correlation of mean $\bar{y}$, and standard deviation $s_y = w$, of samples of five items for Case I. *Mean of $\bar{y}$'s $= \bar{\bar{y}} = 7.141$. Correlation coefficient $r_{w\bar{y}} = -0.020 \pm 0.021$, $s_y = w = 2.079$. Correlation ratio of w on $\bar{y}$, $\eta_{w\bar{y}} = 0.182$ (uncorrected).*

$\bar{y}$

		3.7	4.1	4.5	4.9	5.3	5.7	6.1	6.5	6.9	7.3	7.7	8.1	8.5	8.9	9.3	9.7	10.1	f_w
	4.1								1		2	1							4
	3.9							1	1			1	1						4
	3.7				1				3		1			2					7
	3.5			1					1	1	2	1	1	1					8
	3.3					2	4	2	1		7	2	2	5	1				26
	3.1				1	3	2	2	8	3	9	4	6	3					41
	2.9					5	4	3	10	8	10	3	1	4	2				50
	2.7			1	1		3	12	9	18	18	6	9			4			81
	2.5			2		4	8	14	15	8	22	14	14	6	5	1			113
	2.3				1	2	9	14	10	11	10	10	12	7	3				89
w	2.1			3	4	2	12	12	16	13	15	8	9	7	8	3			112
	1.9	1		1	4	7	6	5	16	7	18	15	7	4	3	4	3		101
	1.7				7	7	9	15	15	23	16	14	17	6	4		2		135
	1.5		1	1	3	5	4	3	6	8	6	7	10	6	5	1	1	1	68
	1.3			2	2	1	3	4	11	8	7	11	3	3	2	2			59
	1.1				1	1	4	5	5	10	6	6	6	6	1			1	52
	0.9			2	1			5	4	2		2	4		1		1		22
	0.7					1		1		7	2	3	2						16
	0.5								3	1	1	1		2		2			10
	0.3									1									1
	0.1									1									1
	f_y	1	1	13	26	40	68	98	135	130	152	109	104	62	35	17	7	2	1000

sufficient condition for the independence of the mean and standard deviation in samples.

Since the number of correlated items, $N = 1000$, is fairly large, we examine into the significance of $\eta_{w\bar{y}} = 0.182$ under the assumption that $N\eta^2_{w\bar{y}}$ is approximately distributed [10] as χ^2 with $a - 1 = 16$ degrees of freedom. This criterion gives odds in favor of significant correlation on approximately a 100 to 1 level of probability.

Next, the means of arrays, $\bar{w}_p$, were plotted to scale on Table 2 to give a general notion of the nature of the regression of $w = s_y$ on $\bar{y}$. The location of these means of arrays of w's affords at least a suggestion of parabolic regression [11] with the curvature concave downward as is to be expected when $\beta_2 - \beta_1 - 3 < 0$, where the β's relate to the parent distribution.

The next step taken was to analyze the variance, as indicated partly in Table 3, where w_i $(i = 1, 2, \cdots, N)$ denotes the stochastic variates, a the number of arrays of w's, $\bar{w}$ the mean of the N values of w_i, $n_p(p = 1, 2, \cdots, a)$ the number of variates in an array marked p, $\bar{w}_p$ the mean of the array marked p, and where the class interval in Table 2, is taken as the unit.

TABLE 3

	Sum of squares	
For deviations of means of arrays of w's	$\sum_{p=1}^{a} n_p(\bar{w}_p - \bar{w})^2 = 380$	$a - 1 = 16$
For deviations of variates from the means of their arrays	$\sum \sum (w_i - \bar{w}_p)^2 = 11{,}098$	$N - a = 983$
Total	$\sum_{i=1}^{N} (w_i - \bar{w})^2 = 11{,}478$	$N - 1 = 999$

In the exhibit given in Table 3, we use the usual algebraic identity

$$\sum_{i=1}^{N} (w_i - \bar{w})^2 = \sum_{p=1}^{a} n_p(\bar{w}_p - \bar{w})^2 + \sum \sum (w_i - \bar{w}_p)^2, \tag{2}$$

where the double sum is made up of a sum of N squares.

By dividing the members of (2) by N, we have

$$\frac{1}{N} \sum_{i=1}^{N} (w_i - \bar{w})^2 = \frac{1}{N} \sum_{p=1}^{a} n_p(\bar{w}_p - \bar{w})^2 + \frac{1}{N} \sum \sum (w_i - \bar{w}_p)^2 \tag{3}$$

The writer has used the identity (3) for many years in lectures to beginners in statistics in proving the equivalence of two definitions of the correlation [12] ratio and is strongly of the opinion that the equality in form (3) appeals more readily to the intuitions of many readers, because of their acquaintance with statements in the language of averages, than does the equivalent equality (2) in the language of sums of squares.

In an extended and more compact form, the analysis is shown in the standard form in Table 4.

TABLE 4

Variance	Degrees of freedom	Sum of squares	Mean square	z-test
Between arrays	16	380	23.75	$\frac{1}{2} \log_e 23.75 = 1.584$
Within arrays	983	11,098	11.29	$\frac{1}{2} \log_e 11.29 = 1.212$
Total	999	11,478		Difference = 0.372

When the sum of squares equal to 380 associated with variance between arrays is further analyzed into a part which could be represented by linear regression,

and a part which represents deviations of the calculated means of arrays of w's from a straight regression line of w on $\bar{y}$, the deviations being measured parallel to the w-axis, we find that the part of the amount 380 represented by linear regression is given by

$$Nr_{w\bar{y}}^2 s_w^2 = 1000\ (.00040)(11.487) = 4.3.$$

Since both $r = .020 \pm 0.021$ and the small value, 4.3, as part of the sum of squares amounting to 380, may well be regarded as sampling fluctuations, we revert to the figures in Table 3 and apply the Fisher z-test. It turns out that the correlation is significant on practically the 100 to 1 level of probability which conforms well with the above inference based on the assumption that $N\eta_{w\bar{y}}^2$ is distributed as χ^2, with $a - 1$ degrees of freedom.

Next, we computed 1000 values of the "Student" ratio $z = (\bar{y} - 7)/w$, for Case I. One of these 1000 values was of the indeterminate form $\frac{0}{0}$. A frequency distribution of the 999 determinate ratios is shown in column (3), Table 5.

By grouping together the class frequencies at the tails of the theoretical distributions until each of the end class frequencies is not less than 5, and calculating χ^2 for the observed distribution in column (3) in comparison with the theoretical distribution in column (6) as found from the "Student" theory in samples of 5 items from a normal distribution, we obtain $\chi^2 = 3.728$ with 11 degrees of freedom.

Thus, the differences between the distribution in column (3) and the "Student" distribution for $N = 5$ shown in column (6) are not only insignificant under the χ^2-test, but are so small as to be expected in a relatively small percentage of statistical experiments even if the "Student" z-distribution were the theoretically exact distribution of our ratios.

The usual moment coefficients of the distribution of observed z's in column (3), Table 5, are:

$$\mu_1' = 0.033533, \qquad \mu_3 = 0.254383, \qquad \beta_1 = 0.55955,$$

$$s = \sqrt{\mu_2} = 0.69799, \qquad \mu_4 = 2.22504, \qquad \beta_2 = 9.37353.$$

Since the value, 0.69799, of the standard deviation of the observed distribution differs very little from $1/\sqrt{N-3} = 0.70711$, the normal curve fitted by using the standard deviation of the observed distribution (column 4, Table 5) differs very little from the normal curve with the origin at the population mean and standard deviation, $\sqrt{2}/2$, (column 5). Furthermore, the application of the χ^2-test to columns (4) and (5) of Table 5 with class frequencies in the "tails" grouped as above gives $\chi^2 = 2.91$ with 9 degrees of freedom.

The moment coefficients of the observed distribution indicate a markedly leptokurtic and somewhat skew distribution but the indications of skewness may be traced mainly and perhaps entirely to the presence of the two extreme variates at the upper end of the distribution and separated about three times the standard deviation from the next class frequency that differs from zero. By

TABLE 5

Distribution of the ratios, $z = (\bar{y} - 7)/w$ in samples of $N = 5$ for Case I.

(1) $z = (\bar{y} - 7)/w$	(2) $t = z\sqrt{N-1} = 2z$	(3) Observed distribution	(4) Normal distribution fitted to observed column (3)	(5) Normal distribution of S.D. $= \frac{1}{\sqrt{N-3}} = \frac{1}{\sqrt{2}}$ in same units as z (measured from population mean)	(6) From the Student theoretical distribution for $N = 5$
−6.0	−12				0.1
−5.5	−11				0.1
−5.0	−10				0.1
−4.5	−9				0.2
−4.0	−8				0.3
−3.5	−7				0.6
−3.0	−6	2	0.05	0.1	1.3
−2.5	−5	1	0.75	0.4	2.7
−2.0	−4	5	3.6	6.2	7.0
−1.5	−3	17	27.7	32.0	21.0
−1.0	−2	67	98.5	105.9	70.5
−0.5	−1	216	210.8	217.2	217.5
0	0	357	279.3	275.4	356.2
0.5	1	226	225.7	217.2	217.5
1.0	2	75	111.7	105.9	70.5
1.5	3	22	33.7	32.0	21.0
2.0	4	5	6.2	6.2	7.0
2.5	5	1	0.75	0.4	2.7
3.0	6	3	0.05	0.1	1.3
3.5	7	0			0.6
4.0	8	0			0.3
4.5	9	0			0.2
5.0	10	2			0.1
5.5	11				0.1
6.0	12				0.1
		999	998.8	999.0	999.0

excluding these two variates from our calculations, we obtain the following moment coefficients:

$$\mu_1' = 0.023571, \qquad \mu_3 = 0.022264, \qquad \beta_1 = 0.0058786,$$

$$s = \sqrt{\mu_2} = 0.662202, \qquad \mu_4 = 1.009673. \qquad \beta_2 = 5.2507062.$$

In the observed distribution thus modified, by excluding the extreme upper class frequency 2, the evidence of skewness has disappeared.

Case II. For our Case II we have a frequency distribution as shown in Table 6.

TABLE 6

Totals in second throws of two dice-values of the stochastic variable....	2	3	4	5	6	7	8	9	10	11	12
Frequency...........................	1	4	9	16	25	36	35	32	27	20	11

Again, since with the uncorrected $\eta_{v\bar{\mu}}$, Table 6, we have $N\eta^2_{v\bar{\mu}} = 31.5$, and since $N\eta^2_{v\bar{\mu}}$ is approximately distributed as χ^2 with $a - 1 = 17$ degree of freedom, we have odds of the order of 100 to 1 against so large a value being a mere sampling fluctuation.

TABLE 7

Correlation of mean $\bar{u}$, and standard deviation $s_u = v$, of five items for Case II, *mean of* $\bar{u} = \bar{\bar{u}} = 6.971$. *Correlation coefficient* $r_{v\bar{u}} = -0.012 \pm 0.020$.
$v = s_u = 2.044$. *Correlation ratio of v on* $\bar{u}$, $\eta_{v\bar{u}} = 0.177$ (*uncorrected*).

$\bar{u}$

v	3.7	4.1	4.5	4.9	5.3	5.7	6.1	6.5	6.9	7.3	7.7	8.1	8.5	8.9	9.3	9.7	10.1	10.5	10.9	f_v
3.9								1		1	1									3
3.7							1				3	1								5
3.5					1	2	1	4	4	2		2								16
3.3						1	5	4	3	5	2	2	1	1	1					25
3.1						4	6	9	6	4	6	1	1	1						38
2.9					3		8	10	16	8	8	5	5		2					65
2.7				1	2	4	10	7	7	17	13	4	1	2	1					69
2.5			1	1	2	10	10	17	8	19	11	5	4	3						91
2.3		1		3	5	5	11	11	14	13	10	7	10	3		1				94
2.1		1	4		3	15	21	16	13	15	13	9	5	3						118
1.9	1		1	1	12	12	7	12	9	16	19	7	13	5						115
1.7		1	1	6	5	11	14	12	16	17	20	10	7	5	2	3				130
1.5				1	7		14	8	8	3	6	5	4	3	3	1	1			64
1.3			1	1	7	10	3	11	3	13	4	1	6	1	1	2				64
1.1		1		2	3	4	5	7	9	9	4	4	5		3	1			1	58
0.9					2	1	2	4	4	6	2	3	1	3	1	1				30
0.7						2	3		1		2	2								10
0.5								2	1	2										5
$f_{\bar{u}}$	1	4	8	16	52	81	121	135	122	150	124	68	63	30	14	9	1	0	1	1000

Now proceeding to the analysis of variance, we substitute our numerical values derived from Table 7 in the identity

$$(4) \qquad \sum_{i=1}^{N} (v_i - \bar{v})^2 = \sum_{p=1}^{a} n_p(\bar{v}_p - \bar{v})^2 + \sum \sum (v_i - \bar{v}_p)^2$$

and obtain, in terms of class intervals as units,

$$10{,}871 = 340 + 10{,}531.$$

An outline of the analysis is exhibited in Table 8

TABLE 8

Variance	Degrees of freedom	Sum of squares	Mean square	z-test
Between arrays.......	17	340	20.00	$\frac{1}{2}\log_e 20.00 = 1.50$
Within arrays.........	982	10,531	10.72	$\frac{1}{2}\log_e 10.72 = 1.18$
Total...............	999	10,871		Diff. $= 0.32$

The moment coefficients and β's which characterize the distribution in Table 6 are:

$$\text{Mean} = 7.972, \quad \mu_2 = 4.888, \quad \mu_3 = -1.755, \quad \mu_4 = 58.724,$$
$$\beta_1 = 0.0264, \quad \beta_2 = 2.449.$$

As in the linear case, samples of 5000 pairs of variates were drawn in sets of five by Miss Chesire. Analogous to Case I, our first concern is with the regression of the standard deviation, $s_u = v$, of u from a sample of five on its mean value, $\bar{u}$.

The correlation table for values of $\bar{u}$ and v is shown in Table 7. The correlation coefficient is
$r_{v\bar{u}} = -0.012 = \pm 0.021$, but the uncorrected correlation ratio of v on $\bar{u}$ is given by

$$\eta_{v\bar{u}} = 0.177.$$

After applying the Pearson and Student corrections, we obtain the corrected

$$\eta_{v\bar{u}} = 0.131.$$

When the sum of squares, 340, associated with variance between arrays is further analyzed into a part which could be represented by linear regression, and a part which represents deviations of the calculated means of arrays of v's from a straight regression line of v on $\bar{u}$, the deviations being measured parallel to the v-axis, we find that the part of the amount 340 represented by linear regression, would be only $Nr^2s_v^2 = 1000\ (.000144)(10.871) = 1.6$.

Since both $r_{v\bar{u}} = -0.012 \pm 0.021$ and the small value, 1.6, as part of the sum of squares 340, may well be regarded as sampling fluctuations, we revert to the figures of Table 8.

The difference of the logarithms in the last column of Table 8, is 0.32, which corresponds to a level of significance of the general order of 100 to 1. Next, we calculate and plot on Table 7 the means of arrays of v's to give a general notion of the regression of v on $\bar{u}$. The location of these means of arrays suggests rather strongly that the regression of v on $\bar{u}$ is parabolic with the curvature concave downward as we should expect from the fact that $\beta_2 - \beta_1 - 3 < 0$, where the β's pertain to the parent distribution.

Next, we computed 1000 values of the "Student" ratio, $z = (u - 7.972)/v$,

for Case II. One of these ratios was infinite. A frequency distribution of the 999 determinate ratios is shown in column 3, Table 9.

The observed distribution (column 3) and the "Student" distribution (column 6) of Table 9, to be expected in samples of $N = 5$, when samples are drawn from a normal distribution, are in close agreement. In fact, when we group together the tail frequencies of the theoretical distribution until each of them is not less than 5, the result of testing the goodness of fit gives $\chi^2 = 17.187$ with 11 degrees of freedom. This gives a value in the neighborhood of 0.1 for the probability, P, that as large or larger deviations than that experienced will occur, due to chance fluctuations, in a single repetition of the experiment. In other words, on the basis of this test, the indications are that we should have in the long run, as large or larger deviations than we have experienced in this case, in about 10 per cent of a large number of sets of sampling of 1000 per set even when the sampling is from a normal distribution.

TABLE 9

Distribution of the ratio, $(\bar{u} - 7.972)/v$ *in samples of five for Case* II.

(1) $z = (\bar{u} - 7.972)/v$	(2) $t = z\sqrt{N-1} = 2z$	(3) Observed	(4) Normal distribution fitted to observed, Column (3).	(5) Normal distribution with S.D. $= \frac{1}{\sqrt{N-3}}$ and origin at population mean	(6) Student's z-distribution for normal parent population with $N = 5$
−5.5	−11	1			0.1
−5.0	−10				0.1
−4.5	−9				0.2
−4.0	−8				0.3
−3.5	−7				0.6
−3.0	−6		0.1	0.1	1.3
−2.5	−5	2	0.4	0.4	2.7
−2.0	−4	3	4.3	6.2	7.0
−1.5	−3	23	25.4	32.0	21.0
−1.0	−2	48	92.0	105.9	70.5
−0.5	−1	203	205.3	217.2	217.5
−0.0	0	380	278.4	275.4	356.2
0.5	1	226	231.4	217.2	217.5
1.0	2	72	117.5	105.9	70.5
1.5	3	24	36.5	32.0	21.0
2.0	4	9	6.9	6.2	7.0
2.5	5	3	0.8	0.4	2.7
3.0	6	4	0.1	0.1	1.3
3.5	7	1			.6
4.0	8				.3
4.5	9				.2
					.1
					.1
					.1
Total		999	999.1	999.0	999.0
∞	∞	1			

SUMMARY

1. The linear correlation coefficient, r, of the mean and standard deviation differs insignificantly from 0 in each case.

2. The correlation ratio of the standard deviation on the mean differs significantly from 0, and the regression of the standard deviation on the mean conforms, in its general aspects, to expectation under the theory of Neyman [12].

3. The indeterminate "Student" ratio of the form, $\frac{0}{0}$, in Case I and that of the form, (constant)/0, in Case II are probably due in part to grouping into class intervals, but the infinite ratio would undoubtedly have had such a large value that it would be excluded from calculations under any one of the known criteria for rejection of extreme observations.

4. Although the rejection of one indeterminate ratio in each of the two cases is slightly disturbing, the evidence presented by our analysis of the experimental sampling lends support to the view that the results of the "Student" theory are almost certainly applicable, for many purposes, when the parent distributions are of such non-normal types as are involved in our sampling.

REFERENCES

[1] W. A. Shewhart and F. W. Winters, "Small samples—new experimental results," *Journal American Statistical Association*, Vol. 23 (1928), pp. 144–153.

[2] P. R. Rider, "On the distribution of the ratio of mean to standard deviation in small samples from non-normal universes," *Biometrika*, Vol. 21 (1929), pp. 124–143. "On small samples from certain non-normal universes," *Annals of Mathematical Statistics*, Vol. 2, (1931), pp. 48–65.

[3] E. S. Pearson, "The distribution of frequency constants in small samples from non-normal symmetrical and skew populations," *Biometrika*, Vol. 21 (1929), pp. 259–286.

[4] M. S. Bartlett, "The effect of non-normality on the *t* distribution," *Proceedings of the Cambridge Philosophical Society*, Vol. 31 (1935), pp. 223–231.

[5] R. C. Geary, "The distribution of "Student's" ratio for non-normal samples," *Supplement to the Journal of the Royal Statistical Society*, No. 2, 1936, pp. 178–184.

[6] H. L. Rietz, "Urn schemeta as a basis for the development of correlation theory," *Annals of Mathematics*, Vol. 21 (1920), pp. 306–322.

[7] E. S. Pearson, Leona Chesire, and Elena M. Oldis," Further experiments on the sampling distribution of the correlation coefficient," *Journal American Statistical Association*, Vol. 27 (1932), pp. 121–128.

[8] Karl Pearson, "On a correction to be made in the correlation ratio," *Biometrika*, Vol. 8, (1911–12), pp. 254–6.

[9] "Student," "The correction to be made in the correlation ratio for grouping," *Biometrika*, (1913), pp. 316–20.

[10] R. A. Fisher, *Statistical methods for research workers*, Fourth Edition, p. 237.

[11] J. Neyman, "On the correlation of mean and variance in samples drawn from an 'infinite' population," *Biometrika*, Vol. 18 (1926), pp. 401–13.

[12] H. L. Rietz, *Mathematical Statistics* (Carus Monograph), 1926, p. 91.

UNIVERSITY OF IOWA, IOWA CITY, IOWA.

THE PROBLEM OF m RANKINGS

By M. G. Kendall and B. Babington Smith

1. **Introduction.** If n objects are ranked by m persons according to some quality of the objects there arises the problem: does the set of m rankings of n show any evidence of community of judgment among the m individuals? For example, if a number of pieces of poetry are ranked by students in order of preference, do the rankings support the supposition that the students have poetical tastes in common, and if so is there any strong degree of unanimity or only a faint degree?

The problem in its full generality permits of no assumption about the nature of the quality according to which the objects are ranked, other than that ranking is possible. No hypothesis is made that the quality is measurable, still less that there is some underlying frequency distribution to the quantiles of which the rankings correspond. The quality is to be thought of as linear in the sense that any two objects possessing it are either coincident or may be put in the relation "before and after." A metric may, of course, be imposed on this linear space by convention; but the relationship between objects is invariant under any transformation which stretches the scale of measurement. In particular, it is not a condition of the problem that the ranking shall be based on a distribution according to a normal variate.

It is permissible to denote the rankings by the *ordinal* numbers 1, 2, $\cdots$ n; but it is not permissible, without further discussion, to operate on these numbers as if they were cardinals. This point seems to have been inadequately appreciated. For instance, when $m = 2$ we have the familiar case of rank correlation between a pair of rankings; and this is frequently treated by subtracting corresponding ranks, squaring, and forming the Spearman coefficient

$$\rho = 1 - \frac{6S(d^2)}{n^3 - n}. \tag{1}$$

To justify this procedure it is necessary to explain what is meant, for example, by such a process as (4th minus 8th), or what the square of this difference of ordinal numbers represents.

It is worth stressing that the necessary transition from ordinals to cardinals can be made without invoking a scale of measurement. When we rank an object as first we mean, in effect, that no member of the set of n is preferred to it; when we rank it as the rth we mean that $(r - 1)$ objects are preferred to it. The ordinals of the ranking are then biunivocally related to the cardinals expressing the number of objects which are preferred. It is thus legitimate

to apply the laws of cardinal arithmetic to them. For example, if an object A is ranked r_1 by Brown, r_2 by Jones and r_3 by Robinson we may form the sum $(r_1 + r_2 + r_3)$, which is to be interpreted as meaning that, taking the preferences of the three persons together, there were $(r_1 + r_2 + r_3 - 3)$ cases in which some other object was preferred to A. The point is of some importance, in view of the prevailing practice of replacing ranks by quantiles of the normal distribution—a practice which cannot always be regarded as justifiable and is sometimes little short of desperate.

To fix the ideas, consider the following three rankings of six objects

Object:	A	B	C	D	E	F
	5	4	1	6	3	2
	2	3	1	5	6	4
	4	1	6	3	2	5
Sum of ranks	11	8	8	14	11	11

We may sum the ranks for each object, as shown. These sums (which must add to 63, and in general to $mn(n + 1)/2$) reflect the degree of resemblance among the rankings. If the resemblance were perfect the sums would be 3, 6, 9, 12, 15, 18 (though not necessarily, of course in that order) and in such a case would be as different as possible. On the other hand, when there is little or no resemblance, as in the example given, the sums are approximately equal. It is thus natural to take the variance of these sums as providing some measure of the concordance in the rankings. If S is the observed sum of squares of the deviations of sums of ranks from the mean value $m(n + 1)/2$ (i.e. is n times the variance) we may write

$$W = \frac{12S}{m^2(n^3 - n)} \tag{2}$$

and call W the coefficient of concordance. Here $m^2(n^3 - n)/12$ is the maximum possible value of S, occurring if there is complete unanimity in the rankings, so that W may vary from 0 to 1. In the example given, $S = 25.5$, $W = 0.16$.

The coefficient W has arisen in several ways.

(a) W is simply related to the average of the $\binom{m}{2}$ Spearman rank correlation coefficients between pairs of the m rankings. It is easy to show that the average ρ is given by

$$\rho_{av} = \frac{\dfrac{12S}{n^3 - n} - m}{m^2 - m} \tag{3}$$

$$= \frac{mW - 1}{m - 1} \tag{4}$$

ρ_{av} has been considered by Kelley [3] as a measure of average intercorrelation in rankings, but he gives no results for testing the significance of observed values.

It is to be noted that whereas W may vary from 0 to 1, ρ_{av} may vary from $-1/(m-1)$ to 1, i.e. it is asymmetrical like the coefficient of intraclass correlation, to which it bears some resemblance.[1]

(b) Friedman [1] has considered a quantity χ_r^2 related to W by the equation

$$\chi_r^2 = m(n-1)W. \tag{5}$$

(c) Welch [6] and Pitman [5] have considered the problem of the distribution of variance in an array

$$\begin{array}{l} a_1, a_2, \cdots a_n \\ b_1, b_2, \cdots b_n \end{array}$$

etc., for permutations of the numbers a, b, etc. in rows.

This is more general than the ranking case, in which $a_1 \cdots a_n$, $b_1 \cdots b_n$ etc. reduce to permutations of the numbers $1 \cdots n$. Since S', the total sum of squares in an array of m rankings of n, is $m^2(n^3 - n)/12$, we have

$$W = \frac{S}{S'} \tag{6}$$

i.e. the ratio of variance between columns to the total variance.

2. **Significance of W.** To test whether an observed value of W is significant it is necessary to consider the distribution of W (or, more conveniently, of S) in the universe observed by permuting the n ranks in all possible ways. No generality is lost by supposing one ranking fixed, and the others will then give rise to $(n!)^{m-1}$ values of S.

The actual distribution of W (or S), as will be seen below, is irregular for low values of m and n, and likely to be quite irregular for moderate values. It may, however, be shown that the first four moments of W are

$$\mu_1' \text{ (about 0)} = \frac{1}{m} \tag{7}$$

$$\mu_2 = \frac{2(m-1)}{m^3(n-1)} \tag{8}$$

$$\mu_3 = \frac{8(m-1)(m-2)}{m^5(n-1)^2} \tag{9}$$

$$\mu_4 = \frac{24(m-1)}{m^7(n-1)^2}\left\{\frac{25n^3 - 38n^2 - 35n + 72}{25(n^3 - n)} + 2(n-1)(m-2) + \frac{n+3}{2}(m-2)(m-3)\right\}. \tag{10}$$

---

[1] The Spearman rank correlation coefficient is the product-moment coefficient of correlation between the ranks considered as ordinary variate values. ρ_{av} is the intraclass correlation coefficient for the m sets of ranks, also considered as variate values.

Results equivalent to these for the first three moments were given by Friedman [1]; and for the first four moments by Pitman [5].

In a valuable contribution to the subject Friedman showed that the distribution of χ_r^2 tends to that of χ^2 with $(n - 1)$ degrees of freedom as m tends to infinity and suggested the use of χ_r^2 (equation (5)) for an ordinary test of significance in the χ^2 distribution. This is satisfactory for moderately large values, but for small values it is subject to the disadvantage inherent in any attempt to represent a distribution of finite range by one of infinite range—the fit near the tails is not likely to be very good. An improvement is obtained by noting that the first four moments of the Type I distribution,

$$df = \frac{1}{B(p, q)} W^{p-1}(1 - W)^{q-1} \tag{11}$$

are approximately those of W if m and n are moderately large, and

$$p = \frac{n-1}{2} - \frac{1}{m} \tag{12}$$

$$q = (m - 1)\left\{\frac{n-1}{2} - \frac{1}{m}\right\}. \tag{13}$$

For practical purposes it is most convenient to put

$$z = \tfrac{1}{2} \log_e \frac{(m-1)W}{1 - W} \tag{14}$$

so that z can be tested in Fisher's distribution with $(n - 1) - \dfrac{2}{m}$ $(= n_1)$ and $(m - 1)\left\{(n - 1) - \dfrac{2}{m}\right\}$ $(= n_2)$ degrees of freedom.

There can be little doubt that this test is quite reliable for moderate values of m and n; but it has hitherto been far from clear how reliable it is for low values of m and n. This point we attempt to clear up in the present paper.

3. **Distribution of S.** For the case $m = 2$ the distribution of S is the same as the distribution of the $S(d^2)$ used in calculating Spearman's rank correlation coefficient. A table showing the distribution up to and including $n = 8$ has already been given (Kendall and others, [4]). Tables giving probabilities that specified values of χ_r^2 would be attained or exceeded were given by Friedman [1] for $n = 3$, $m = 2$–9; and $n = 4$, $m = 2$–4. We have taken this work somewhat further and obtained the distributions for $n = 3$, $m = 2$–10; $n = 4$, $m = 2$–6; and $n = 5$, $m = 3$. Tables 1–4 give the probabilities based on these distributions.

These distributions were obtained by two methods. The first consisted of building up the distribution for $(m + 1)$ and n from that of m and n. For

TABLE 1

Probability that a given value of S will be attained or exceeded, for $n = 3$ and values of m from 2 to 10

Values of m

S	2	3	4	5	6	7	8	9	10
0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
2	.866	.944	.931	.954	.956	.964	.967	.971	.974
6	.500	.528	.653	.691	.740	.768	.794	.814	.830
8	.167	.361	.431	.522	.570	.620	.654	.685	.710
14		.194	.273	.367	.430	.486	.531	.569	.601
18		.028	.125	.182	.252	.305	.355	.398	.436
24			.069	.124	.184	.237	.285	.328	.368
26			.042	.093	.142	.192	.236	.278	.316
32			.0046	.039	.072	.112	.149	.187	.222
38				.024	.052	.085	.120	.154	.187
42				.0085	.029	.051	.079	.107	.135
50				$.0^{3}77$	.012	.027	.047	.069	.092
54					.0081	.021	.038	.057	.078
56					.0055	.016	.030	.048	.066
62					.0017	.0084	.018	.031	.046
72					$.0^{3}13$	.0036	.0099	.019	.030
74						.0027	.0080	.016	.026
78						.0012	.0048	.010	.018
86						$.0^{3}32$	.0024	.0060	.012
96						$.0^{3}32$	.0011	.0035	.0075
98						$.0^{4}21$	$.0^{3}86$	.0029	.0063
104							$.0^{3}26$	.0013	.0034
114							$.0^{4}61$	$.0^{3}66$	.0020
122							$.0^{4}61$	$.0^{3}35$	.0013
126							$.0^{4}61$	$.0^{3}20$	$.0^{3}83$
128							$.0^{5}36$	$.0^{4}97$	$.0^{3}51$
134								$.0^{4}54$	$.0^{3}37$
146								$.0^{4}11$	$.0^{3}18$
150								$.0^{4}11$	$.0^{3}11$
152								$.0^{4}11$	$.0^{4}85$
158								$.0^{4}11$	$.0^{4}44$
162								$.0^{6}60$	$.0^{4}20$
168									$.0^{4}11$
182									$.0^{5}21$
200									$.0^{7}99$

TABLE 2

Probability that a given value of S will be attained or exceeded for $n = 4$ and $m = 3$ and 5

S	$m = 3$	$m = 5$	S	$m = 5$
1	1.000	1.000	61	.055
3	.958	.975	65	.044
5	.910	.944	67	.034
9	.727	.857	69	.031
11	.608	.771	73	.023
13	.524	.709	75	.020
17	.446	.652	77	.017
19	.342	.561	81	.012
21	.300	.521	83	.0087
25	.207	.445	85	.0067
27	.175	.408	89	.0055
29	.148	.372	91	.0031
33	.075	.298	93	.0023
35	.054	.260	97	.0018
37	.033	.226	99	.0016
41	.017	.210	101	.0014
43	.0017	.162	105	$.0^364$
45	.0017	.141	107	$.0^333$
49		.123	109	$.0^321$
51		.107	113	$.0^314$
53		.093	117	$.0^448$
57		.075	125	$.0^530$
59		.067		

example, with $m = 2$ and $n = 3$ we have the following values of the sums of ranks, measured about their mean:

Type			Frequency
-2	0	2	1
-2	1	1	2
-1	0	1	2
0	0	0	1

Here $-2, 1, 1$, and $2, -1, -1$ are taken to be identical types, for they give the same value of S and will also give similar types when we proceed to $m = 3$ as follows.

In the case $m = 3$ each of the above type will appear added to the six permutations of $-1, 0, 1$; e.g. the type $-2, 0, 2$ will give one each of $-3, 0, 3$; $-3, 1, 2$;

TABLE 3

Probability that a given value of S will be attained or exceeded for $n = 4$ and $m = 2, 4$ and 6

S	$m = 2$	$m = 4$	$m = 6$	S	$m = 6$
0	1.000	1.000	1.000	82	.035
2	.958	.992	.996	84	.032
4	.833	.928	.957	86	.029
6	.792	.900	.940	88	.023
8	.625	.800	.874	90	.022
10	.542	.754	.844	94	.017
12	.458	.677	.789	96	.014
14	.375	.649	.772	98	.013
16	.208	.524	.679	100	.010
18	.167	.508	.668	102	.0096
20	.042	.432	.609	104	.0085
22		.389	.574	106	.0073
24		.355	.541	108	.0061
26		.324	.512	110	.0057
30		.242	.431	114	.0040
32		.200	.386	116	.0033
34		.190	.375	118	.0028
36		.158	.338	120	.0023
38		.141	.317	122	.0020
40		.105	.270	126	.0015
42		.094	.256	128	$.0^{3}90$
44		.077	.230	130	$.0^{3}87$
46		.068	.218	132	$.0^{3}73$
48		.054	.197	134	$.0^{3}65$
50		.052	.194	136	$.0^{3}40$
52		.036	.163	138	$.0^{3}36$
54		.033	.155	140	$.0^{3}28$
56		.019	.127	144	$.0^{3}24$
58		.014	.114	146	$.0^{3}22$
62		.012	.108	148	$.0^{3}12$
64		.0069	.089	150	$.0^{4}95$
66		.0062	.088	152	$.0^{4}62$
68		.0027	.073	154	$.0^{4}46$
70		.0027	.066	158	$.0^{4}24$
72		.0016	.060	160	$.0^{4}16$
74		$.0^{3}94$	.056	162	$.0^{4}12$
76		$.0^{3}94$	.043	164	$.0^{5}80$
78		$.0^{3}94$	.041	170	$.0^{5}24$
80		$.0^{4}72$	.037	180	$.0^{6}13$

TABLE 4

Probability that a given value of S will be attained or exceeded, for $n = 5$ and $m = 3$

S	$m = 3$	S	$m = 3$
0	1.000	44	.236
2	1.000	46	.213
4	.988	48	.172
6	.972	50	.163
8	.941	52	.127
10	.914	54	.117
12	.845	56	.096
14	.831	58	.080
16	.768	60	.063
18	.720	62	.056
20	.682	64	.045
22	.649	66	.038
24	.595	68	.028
26	.559	70	.026
28	.493	72	.017
30	.475	74	.015
32	.432	76	.0078
34	.406	78	.0053
36	.347	80	.0040
38	.326	82	.0028
40	.291	86	$.0^390$
42	.253	90	$.0^469$

$-2, -1, 3; -2, 1, 1; -1, -1, 2;$ and $-1, 0, 1$. These types are then counted for each of the four basic types of $m = 2$ and we get:

Type			Frequency
−3	0	3	1
−3	1	2	6
−2	0	2	6
−2	1	1	6
−1	0	1	15
0	0	0	2

The case $m = 4$ is treated by considering the numbers of types obtained by adding the six permutations of $-1, 0, 1$ to the types for $m = 3$; and so on.

This method is quite convenient for $n = 2$ and $n = 3$. For $n = 4$ it becomes difficult owing to the labour of considering 24 permutations at each stage and to the increase in the number of types. For $n = 5$ there are 120 permutations and the labour becomes excessive.

The second method employed is a generalisation of a procedure we used for the Spearman coefficient. Taking first of all the case $m = 2$, consider the array

$$\begin{array}{cccc} a^2 & a^3 & a^4 & \cdots a^{(n+1)} \\ a^3 & a^4 & a^5 & \cdots a^{(n+2)} \\ \cdots & \cdots & \cdots & \cdots \\ a^{(n+1)} & a^{(n+2)} & a^{(n+3)} & \cdots a^{2n} \end{array}$$

Any permissible set of values of the sums of ranks is obtained by selecting n entries from this array so that no entry appears more than once in the same row or column. If then, subtracting from each index the mean $(n + 1)$ and squaring, we write

$$E = \left\{ \begin{array}{ccccc} a^{(n-1)^2} & a^{(n-2)^2} & \cdots & a^1 & a^0 \\ a^{(n-2)^2} & a^{(n-3)^2} & \cdots & a^0 & a^1 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a^0 & a^1 & \cdots & a^{(n-2)^2} & a^{(n-1)^2} \end{array} \right\} \tag{15}$$

the values of S are the powers of a in E when it is expanded as a sum of $n!$ terms each of which is obtained by multiplying n factors which do not appear in the same row or column. The distribution of S is arrayed by the expansion of E, the number of values of any S being the coefficient of a^S in the expansion.

Similarly, for m rankings, the distribution of S is given by the expansion of an m- dimensional E-function. For example, with $m = 3$ there would be a three-dimensional E-function the bottom plane of which would be

$$\begin{array}{cccc} a^{\left\{3-\frac{3(n+1)}{2}\right\}^2} & a^{\left\{4-\frac{3(n+1)}{2}\right\}^2} & \cdots & a^{\left\{n+2-\frac{3(n+1)}{2}\right\}^2} \\ a^{\left\{4-\frac{3(n+1)}{2}\right\}^2} & a^{\left\{5-\frac{3(n+1)}{2}\right\}^2} & \cdots & a^{\left\{n+3-\frac{3(n+1)}{2}\right\}^2} \\ \cdots & \cdots & \cdots & \cdots \\ a^{\left\{n+2-\frac{3(n+1)}{2}\right\}^2} & a^{\left\{n+3-\frac{3(n+1)}{2}\right\}^2} & \cdots & a^{\left\{2n+2-\frac{3(n+1)}{2}\right\}^2} \end{array}$$

The plane above this would be

$$\begin{array}{ccc} a^{\left\{4-\frac{3(n+1)}{2}\right\}^2} & \cdots & a^{\left\{n+3-\frac{3(n+1)}{2}\right\}^2} \\ \cdots & \cdots & \cdots \\ a^{\left\{n+3-\frac{3(n+1)}{2}\right\}^2} & \cdots & a^{\left\{2n+3-\frac{3(n+1)}{2}\right\}^2} \end{array}$$

and so on.

The E-function is difficult to handle in more than three dimensions, but for the two and three dimensional case it is manageable and we used it to obtain the distribution of S for $n = 5$ and $m = 3$.

4. **Adequacy of the z-test.** Tables 1–4 provide exact tests for the values of m and n there given. It remains to be seen how good the ordinary z-test applied to W would be for higher values. It may be presumed that if the test is satis-

factory for any particular value of m and n for which exact results are available, it will be so for higher values of m and n. Since, for ordinary purposes the significance points of z as tabled by Fisher and Yates [2] would be employed, the most useful comparison would seem to be between those tables and the extreme values of tables 1–4.

For $n = 3$, $m = 9$, the 1% level is given approximately by $S = 78$. We have, testing for such a value, $W = 0.4814$, $z = 1.002$, $n_1 = \frac{16}{9}$, $n_2 = \frac{128}{9}$. By linear interpolation of reciprocals in the tables of z we find for the 1% point and these degrees of freedom $z = 0.954$. The correspondence is hardly satisfactory, and the z test might lead to incorrect inferences in practice. Matters improve a good deal, however, if we make continuity corrections, by subtracting unity from S before calculating W, and increasing by two the divisor $m^2(n^3 - n)/12$, so as to allow for the finite range. In this case $z = 0.979$.

For $n = 4$, $m = 6$ the 1% point is approximately $S = 100$. We have $W = 0.5556$, $z = 0.916$, $n_1 = 8/3$, $n_2 = 40/3$. By linear interpolation as before we find $z = 0.888$.

Continuity corrections again materially improve the agreement, giving a value of $z = 0.893$.

For $n = 5$ $m = 3$ there is no very convenient value of S close to the 1% point. For $P = 0.015$ $S = 74$ and for $P = 0.0078$ $S = 76$.

For $S = 74$ (with continuity corrections) $z = 1.020$
$S = 76$ (" " ") $z = 1.089$

By interpolation from the tables $z = 1.075$. The use of the z test would lead to the correct conclusion that a value of S equal to 74 falls below, and that of 76 above, the 1% point.

For values of m and n not included in Tables 1–4 it thus appears that the z-test with continuity corrections will give sufficiently accurate results, if n is greater than 3, at the 1% points. It may be presumed that the results at the 5% points are equally good and probably better. But for finer values of significance, such as 0.1%, it is doubtful whether the test is sound. The tails of the distribution of S for moderate values of m and n are very irregular.

For instance, the following is the tail of the distribution of S for $n = 3, m = 10$ (the total distribution being 10,077,696):

S	Frequency	S	Frequency
96	11,340	146	740
98	30,090	150	252
104	13,830	152	420
114	7,380	158	240
122	4,200	162	90
126	3,240	168	90
128	1,450	182	20
134	1,860	200	1

and the following is the tail for $n = 4$ $m = 6$ (the total being 7,962,624):

S	Frequency	S	Frequency	S	Frequency
100	5536	122	4100	146	810
102	8160	126	4480	148	225
104	10260	128	240	150	264
106	8850	130	1152	152	120
108	3920	132	660	154	180
110	13344	134	1980	158	60
114	5460	136	300	160	36
116	3870	138	600	162	30
118	3900	140	312	164	45
120	2472	144	100	170	18
				180	1

Irregularities of this kind run all through the distributions we have obtained, and frequency diagrams present the same sort of features we have noticed in the case $m = 2$ (Kendall and others, [4]). The representation of such distributions by continuous functions, no matter how close their lower moments, is obviously to be used with some care. Although the B-distribution or the associated z-distribution will give reasonable significance tests at levels of 1% or greater, they will probably be inadequate to represent frequencies occurring in narrow ranges.

5. **Some Experimental Distributions.** In some previous work we obtained a number of random permutations of the numbers 1–10 and 1–20. These were used to derive some experimental distributions of S which may be worth recording. Table 5 gives the distribution for 200 sets of pentads of 10 and Table 6 that for 100 triads of 20. In the distribution of Table 5, the mean of the grouped distribution is 404. The theoretical mean is 412.5 with a standard error of 12.3. In Table 6 the mean is 1936, the theoretical mean being 1995 with s.e. 53. The distributions accord quite well with expectation.

In conclusion we give two examples to illustrate some points of importance in ranking work. The first is a case in which ranks appear as the primary variate and in which the assumption of normality is clearly illegitimate.

6. **Example 1.** In some experiments in random series a pack of ordinary playing cards was shuffled and the order of the 13 cards of each suit from the top of the pack was noted. The pack was then reshuffled and again the orders noted. This was done 28 times. The question we wished to discuss was whether the shuffling was good, in the sense that the cards were thoroughly mixed at each shufflle.

Here, for each suit, say diamonds, we have 28 rankings of 13. The sums of ranks were 183, 137, 171, 207, 188, 160, 225, 174, 216, 192, 236, 239, 220. The mean is 196, and $S = 11522$, W (without continuity corrections, which are not

TABLE 5

Experimental Distribution of S in 200 sets ($m = 5$, $n = 10$)

S	Frequency
0–	1
50–	2
100–	7
150–	9
200–	21
250–	22
300–	24
350–	26
400–	20
450–	17
500–	12
550–	11
600–	10
650–	4
700–	5
750–	3
800–	3
...	..
1000–	2
...	..
1250–	1
Total	200

TABLE 6

Experimental Distribution of S in 100 sets ($m = 3$, $n = 20$)

S	Frequency
800–	4
1000–	8
1200–	8
1400–	6
1600–	12
1800–	15
2000–	20
2200–	12
2400–	6
2600–	5
2800–	0
3000–	3
3200–	0
3400–	1
Total	100

worth making for these values of m and n) = 0.08075, z (equation (14)) = 0.432. This falls just beyond the 1% point.

Similarly for the clubs W was found to be 0.0535; for the hearts, 0.0245; and for spades, 0.0342. None of these values is significant and we conclude that the randomisation introduced by the shuffling was good, at all events, so far as this test was concerned. It may be added that the shuffling was done with much more care than would be taken in an ordinary game of cards.

In psychological work there has sometimes been a confusion between the determination of a measure of agreement between subjects and that of an objective order based on experimental rankings. It may therefore be as well to point out that in its psychological applications the test of W is one of concordance between judgments. There may be quite a high measure of agreement about something which is incorrect.

7. **Example 2.** A number of students were given 12 photographs of persons unknown to them, and asked to rank them in what they judged from the photographs to be their intelligence. For 16 students the sums of ranks were

112, 94, 101, 84, 97, 75, 104, 84, 102, 146, 125, 124

The mean is 104. $S = 4472$, $W = 0.1222$. $z = 0.368$, and is barely significant, being between the 1% and the 5% points.

For 111 students the sums were

818, 670, 908, 410, 706, 526, 780, 485, 596, 1044, 959, 756

$$W = 0.2378, z = 1.768$$

This is highly significant and it is to be inferred that community of judgment exists between students or groups of students. But there was little relationship between the judgments and the intelligence of the photographed subjects as given by the Binet Intelligence Quotient.

Note added in proof:

While this paper was passing through the press Professor W. Allen Wallis, of Stanford University, kindly drew our attention to some unpublished work of his own on this subject. Professor Wallis had also arrived at the coefficient W which, he points out, is the ranking analogue of the correlation ratio. His paper is, we understand, on the point of publication.

REFERENCES

[1] M. Friedman, "The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance." *Jour. Am. Stat. Assn.*, Vol. 32(1937), p. 675.

[2] R. A. Fisher, and F. Yates, *Statistical Tables for Biological, Agricultural and Medical Research*, 1938, Oliver & Boyd, Edinburgh.

[3] T. L. Kelley, *Statistical Methods*, 1927.

[4] M. G. Kendall, Sheila F. H. Kendall, and Bernard Babington Smith, "The Distribution of Spearman's Coefficient of Rank Correlation in a Universe in which all Rankings Occur an Equal Number of Times," *Biometrika*, Vol. 30(1938), p. 251.

[5] E. J. G. Pitman, "Significance Tests which may be applied to Samples from any Populations: III. The analysis of variance." *Biometrika*, Vol. 29(1938), p. 322.

[6] B. L. Welch, "On the z- test in Randomised Blocks and Latin Squares," *Biometrika*, Vol. 29(1937), p. 21.

LONDON, ENGLAND
AND
UNIVERSITY OF ST. ANDREWS,
SCOTLAND.

NOTES

This section is devoted to brief research and expository articles, notes on methodology and other short items.

THE ALLOCATION OF SAMPLINGS AMONG SEVERAL STRATA

BY J. STEVENS STOCK AND LESTER R. FRANKEL

1. Introduction. The problem of selecting a random sample so as to obtain optimum precision in making estimates has been the subject of inquiries by Bowley,[1] Neyman,[2] Sukhatme[3] and others. In estimating an average value of a variate in a population it is often profitable to stratify the universe into several homogeneous parts and sample at random within each of these parts. In order to obtain maximum efficiency for a given size of sample it appears that the number of samplings from each stratum should be proportional to the standard deviation of the characteristic under consideration and to the total number of units within the stratum. By distributing the sample in such a manner optimum precision will be obtained in estimating a general average.

However, it often happens that it is not the purpose of an investigation to study the aggregate of the universe. Evaluations and interrelations of characteristics in different groups or strata within the universe may be of importance. Thus, in cost-of-living surveys in a number of urban centers the object is to compare costs among the cities of different backgrounds. In such cases it is desirable for each city to have equal reliability so that each one may be treated as a unit. There are many other situations in the social sciences where analyses of this type are of importance.

2. The Problem. In general, the sampling problem is: Given several well defined areas of study and a fixed number of observations with which to make the survey, how best to distribute the observations such that each area will be represented with equal precision.

There are n observations to be distributed among m areas or strata. In the

---

[1] A. L. Bowley, "Measurement of the precision attained in sampling," *Bulletin de l'Institute International de Statistique* 1926 Rome, Tome XXII, 1-ere Livraison, 3-eme partie, pp. 1–62 (supplement).

[2] J. Neyman, "On the two different aspects of the representative method," *Journal of the Royal Statistical Society*, 1934, pp. 558–625.

[3] P. V. Sukhatme, "Contribution to the theory of the representative method," Supplement to the *Journal of the Royal Statistical Society*, Vol. II, 1935, pp. 253–268.

i-th stratum, if N_i is the total number of units, S_i^2 the variance of the characteristic to be measured, and n_i the size of the sample, the sampling error of the arithmetic mean is

$$\sigma_i = \sqrt{\frac{S_i^2}{n_i}\frac{(N_i - n_i)}{(N_i - 1)}}. \tag{1}$$

The problem then is, given N_i, numbers proportional to S_i^2 and n, to determine n_i such that

$$\sigma_1 = \sigma_2 = \cdots = \sigma_m.$$

3. **First Solution.** If we assume that the variances S_i^2 are all equal and that for $N_i - 1$ we may substitute N_j, we have

$$\frac{N_1 - n_1}{n_1 N_1} = \frac{N_2 - n_2}{n_2 N_2} = \cdots = \frac{N_m - n_m}{n_m N_m}. \tag{2}$$

From the total amount of money available and the cost per sampling unit we can determine the total number of observations to be apportioned among the m populations

$$n = \sum_1^m n_i. \tag{3}$$

We are able then to write m equations in m unknowns:
From (2) we may write $m - 1$ equations

$$\frac{1}{n_1} - \frac{1}{N_1} = \frac{1}{n_j} - \frac{1}{N_j} \qquad (j = 2, 3, \cdots m) \tag{4}$$

and from (3) we may write one equation.

$$n_1 + n_2 + \cdots + n_m = n. \tag{5}$$

But equations (4) are not easily soluble in their present form; they can be made linear by writing the approximation

$$\frac{1}{n_i} \equiv \frac{1}{L_i(1 + \alpha_i)} \doteq \frac{1 - \alpha_i}{i_i}.$$

Where L_i is some reasonable approximation of n_i chosen such that

$$\sum_1^m L_i = \sum_1^m n_i$$

and α_i is some small correction for L_i to be determined. We have then approximately,

$$\frac{1 - \alpha_1}{L_1} - \frac{1}{N_1} = \frac{1 - \alpha_j}{L_j} - \frac{1}{N_j} \qquad (j = 2, 3, \cdots m) \tag{6}$$

and from equation (5) we get

$$\alpha_1 L_1 + \alpha_2 L_2 + \cdots + \alpha_m L_m = 0. \tag{7}$$

If we write

$$\phi_i \equiv L_1 L_i\left(\frac{1}{N_1} - \frac{1}{N_i}\right) + L_1 - L_i$$

we may write (6) and (7) in the following form:

$$(8)\qquad \begin{array}{l} -L_2\alpha_1 + L_1\alpha_2 \hspace{12em} = \phi_1 \\ -L_3\alpha_1 \hspace{3em} + L_1\alpha_3 \hspace{7em} = \phi_3 \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ -L_m\alpha_1 \hspace{10em} + L_m\alpha_1 = \phi_m \\ \quad L_1\alpha_1 + L_2\alpha_2 + L_3\alpha_3 + \cdots + L_m\alpha_m = 0 \end{array}$$

The matrix of the coefficients is

$$(9)\qquad \left\| \begin{array}{ccccc} -L_2 & L_1 & 0 & \cdots & 0 \\ -L_3 & 0 & L_1 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ -L_m & \cdot & \cdot & \cdots & L_1 \\ L_1 & L_2 & \cdot & \cdots & L_m \end{array} \right\|$$

From this matrix we find that

$$(10)\qquad \alpha_1 = \frac{-\sum_2^m \phi_i L_i}{\sum_2^m L_i^2}$$

and from the general form of equation (8) we have

$$(11)\qquad \alpha_i = \frac{\phi_i + L_i\alpha_1}{L_1}.$$

These two equations (10) and (11) give us all the α_i. It is then only necessary to compute the second approximations of n_i by

$$(12)\qquad L_i' = L_i(1 + \alpha_i) \doteq n_i.$$

Closer approximations, though perhaps unnecessary, can be made by repeating the computation with the next approximations. The final approximations may be checked by substituting them in equations (4).

4. **Second Solution.** Sometimes the numbers S_i^2 are known or at least proportionate numbers can be estimated with a fair degree of accuracy for each area. We shall call these proportionate number ξ_i^2. We now have the conditions

$$(13)\qquad \xi_1^2\frac{N_1 - n_1}{n_1N_1} = \xi_2^2\frac{N_2 - n_2}{n_2N_2} = \cdots = \xi_m^2\frac{N_m - n_m}{n_mN_m}$$

and as before

$$\sum_1^m n_i = n .$$

We may write m equations in m unknowns, α_i, using the approximations L_i as before:[4]

$$(14)\qquad \begin{array}{lllll} -\,S_1^2L_2\alpha_1 & +\,S_2^2L_1\alpha_2 & & & = \theta_2 \\ -\,S_1^2L_2\alpha_1\,\cdot & & +\,S_3^2L_1\alpha_3 & & = \theta_3 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ -\,S_1^2L_m\alpha_1 & & & +\,S_m^2L_1\alpha_m & = \theta_m \\ \quad L_1\alpha_1 & +\,L_2\alpha_2 & +\,\cdots & +\,L_m\alpha_m & = 0 \end{array}$$

Where

$$(15)\qquad \theta_i = L_1S_i^2 - L_iS_i^2 + L_iL_1\left(\frac{S_1^2}{N_1} - \frac{S_i^2}{N_i}\right).$$

Solving these m linear equations for α_i we get

$$\alpha_1 = \frac{-\sum_2^m \theta_i\, L_i/S_i^2}{S_1^2 \sum_2^m L_i^2/S_i^2}$$

and from the general form of equations (14) we have

$$\alpha_i = \frac{\theta_i + S_1^2 L_i\alpha_1}{S_i^2 L_1} .$$

These α_i may be applied as before to the approximations L_i for new approximations L_i' of the numbers n_i.

5. **Remarks.** (i) In either case the applications of the corrections to the approximation L_i may be applied in two different ways:

$$(16)\qquad L_i' = L_i(1 + \alpha_i)$$

$$(17)\qquad L_i' = \frac{L_i}{1 - \alpha_i}.$$

When the corrections are applied according to (16) the sum of the new approximations adds up correctly to the total n, and no further adjustment need be made in the L_i' either for repeating the operation again for nearer approximations or for final use. If, however, the corrections are relatively large, say

[4] The numbers S_i^2 and ξ_i^2 may be used interchangeably since they are by hypothesis proportional.

greater than .10, there seems to be better convergence with the second approximations if formula (17) is used and the resulting L_i' adjusted proportionately such that they add up to n. These numbers then can again be adjusted with new α_i for final approximations.

(ii) The numbers S_i^2 or ξ_i^2 are not always estimable. If they are not known at all or are known to be all nearly equal the first solution is perhaps the more useful. If these numbers are known, and known to be different, the second solution is necessary. However, some saving in computation by the second method may be effected if the approximations L_i are first adjusted by the first solution before being entered into the computation of the second solution.

(iii) Further accuracy, though perhaps unnecessary, may be attained in the second solution by substituting throughout $S_i'^2$ for S_i^2 where

$$S_i'^2 = \frac{N_i}{N_i - 1} S_i^2$$

This substitution eliminates any slight inaccuracies caused by substituting N_i for $N_i - 1$.

(iv) The initial approximations L_i may in almost every case be gotten from the following formula:

$$L_i = \frac{n}{m} - \left(\frac{n}{m}\right)^2 \left(\frac{1}{N_i} - \frac{1}{m} \sum_1^m \frac{1}{N_i}\right).$$

(v) In all that has been presented above it has been assumed that the sample has been drawn without replacements from a finite universe. Whether or not this assumption is tenable depends upon the particular object of the research.

6. **Example.** In the Survey of Youth in the Labor Market conducted by the Division of Research in the Works Progress Administration youth who completed the eighth grade in the school years 1928–1929, 1930–1931 and 1932–1933 were studied. In six cities, Duluth, Denver, Birmingham, Seattle, San Francisco, and St. Louis random samples from school records were selected. Funds permitted the use of 40,000 schedules.

From school records it was possible to determine the total number of eighth grade graduates in each city for the years in question. The problem arose then as to what would be the most efficient method of distributing these 40,000 schedules among the six cities in order to compare the problems of youth.

Assuming equal variances within cities, quotas were computed for each of the cities. From Table 1, summarizing the computations, it can be seen that the quotas fall somewhere between proportionate and equal frequencies. This last result would be expected if samplings had been made from infinite universes.

7. **Note.** In the social sciences interest centers in deriving relationships among the various strata where each stratum is considered as a single unit. In such cases equal precision is desired. However, if the object of research is

TABLE 1

City	8th grade graduates	Initial approximation	First correction term	First approximation	Second correction term	Quotas	Percent sampled
Duluth, Minn.	5,500	4,000	−.02968	3,881	−.00077	3,878	70.51
Birmingham, Ala.	9,000	5,500	+.06641	6,399−	+.00148	5,343	59.37
Denver, Colo.	12,500	6,000	−.02690	5,352	−.00164	6,409	51.27
Seattle, Wash.	15,000	6,500	+.07525	6,989	+.00257	7,007	46.71
San Francisco, Cal.	21,000	8,000	+.01425	8,114	−.00341	8,086	38.50
St. Louis, Mo.	31,000	10,000	−.07349	9,265	+.00129	9,277	29.93
Total	94,000	40,000		40,000		40,000	

simply to draw contrasts between any two strata we would seek to minimize the standard error of the difference,

$$\sigma_{\Delta_{jk}} = \sqrt{S_j'^2\left(\frac{1}{n_j} - \frac{1}{N_j}\right) + S_k'^2\left(\frac{1}{n_k} - \frac{1}{N_k}\right)}$$

subject to the condition,

$$\sum_1^m n_i = n\,.$$

This leads to the result

$$\frac{S_j'}{n_j} = \frac{S_k'}{n_k}.$$

Thus, the number of samplings from each stratum is, for all practical purposes, proportional to the standard deviations, irrespective of the size of the various strata.

WASHINGTON, D. C.

ON THE COEFFICIENTS OF THE EXPANSION OF $X^{(n)}$

By J. A. Joseph

Let us construct the following triangular arrangement of numbers:

$$\begin{array}{c}
1 \\
1 \quad\quad 1 \\
1 \quad\quad 3 \quad\quad 2 \\
1 \quad\quad 6 \quad\quad 11 \quad\quad 6 \\
1 \quad\quad 10 \quad\quad 35 \quad\quad 50 \quad\quad 24 \\
\cdot \quad\quad \cdot \quad\quad \cdot \quad\quad \cdot \quad\quad \cdot \quad\quad \cdot \\
1 \quad f_1(n-1) \quad f_2(n-1) \quad \cdot \quad \cdot \quad \cdot \quad f_{n-2}(n-1) \quad f_{n-1}(n-1) \\
1 \quad f_1(n) \quad f_2(n) \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad f_{n-1}(n) \quad f_n(n)
\end{array}$$

where the n-th row can be constructed from the preceding row by means of the expression

$$n \cdot f_i(n-1) + f_{i+1}(n-1) = f_{i+1}(n). \tag{1}$$

For example, the element 35 in the middle of the 4th row is obtained from the two elements immediately above it, $4 \cdot 6 + 11 = 35$. (The top element is counted as the zeroth row.)

The elements in the $(n-1)$st row are the coefficients in the expansion of $x^{(n)}$ as a function of x, using the notation of the calculus of finite differences. For example,

$$\begin{aligned} x^{(4)} &= x(x-1)(x-2)(x-3) \\ &= x^4 - 6x^3 + 11x^2 - 6x. \end{aligned}$$

Of course, the signs of the coefficients alternate.

The function $f_i(n)$ is the sum of the products of the first n integers taken i at a time, namely

$$f_i(n) = \sum^{n} \epsilon_1 \epsilon_2 \cdots \epsilon_i \tag{2}$$

the summation being a symmetric function of the integers $1, 2, 3, \cdots, n$.

Equation (1) can be written as a linear, first order difference equation,

$$\begin{aligned} \Delta f_{i+1}(n-1) &\equiv f_{i+1}(n) - f_{i+1}(n-1) = n \cdot f_i(n-1) \\ f_{i+1}(n-1) &= \Delta^{-1}[n \cdot f_i(n-1)]. \end{aligned} \tag{3}$$

Since $f_0(n) = 1$ for all values of n, we can find $f_1(n)$, and consequently $f_2(n)$, and so on. Thus

$$\begin{aligned} f_1(n-1) &= \Delta^{-1} n = \frac{n^{(2)}}{2} \\ f_2(n-1) &= \Delta^{-1}\left[n \cdot \frac{n^{(2)}}{2}\right] = \frac{3n^{(4)} + 8n^{(3)}}{24} \\ f_3(n-1) &= \Delta^{-1}\left[n\left(\frac{3n^{(4)} + 8n^{(3)}}{24}\right)\right] \\ &= \frac{n^{(6)} + 8n^{(5)} + 12n^{(4)}}{48}. \end{aligned} \tag{4}$$

The following theorems are true for the "triangle":

THEOREM 1. *The sum of the elements in any* n-th *row is equal to* $(n+1)!$, *namely,*

$$\sum_{i=0}^{i} f_i(n) = (n+1)! \tag{5}$$

THEOREM 2. *The sum of the even elements of any row is equal to the sum of the odd elements, or*

$$\sum_{i=0}^{n} (-1)^i f_i(n) = 0. \tag{6}$$

From these coefficients we can generate the Bernoulli numbers:

$$
\begin{aligned}
&B_0 = \frac{1}{2}\\
&B_0 - B_1 = \frac{2!}{3}\\
&2B_0 - 3B_1 + B_2 = \frac{3!}{4}\\
(7)\quad &6B_0 - 11B_1 + 6B_2 - B_3 = \frac{4!}{5}\\
&24B_0 - 50B_1 + 35B_2 - 10B_3 + B_4 = \frac{5!}{6}\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\
&f_n(n)B_0 - f_{n-1}(n)B_1 + f_{n-2}(n)B_2 - \cdots + (-1)^n f_0 B_n = \frac{(n+1)!}{n+2}
\end{aligned}
$$

Or, as a determinant:

$$
|B_n| = \begin{vmatrix}
\frac{1}{2} & 1 & 0 & 0 & \cdots & 0\\
\frac{2!}{3} & 1 & 1 & 0 & \cdots & 0\\
\frac{3!}{4} & 2 & 3 & 1 & \cdots & 0\\
\frac{4!}{5} & 6 & 11 & 6 & \cdots & 0\\
\vdots & \vdots & \vdots & \vdots & & \vdots\\
\frac{(n+1)!}{n+2} & f_n(n) & f_{n-1}(n) & f_{n-2}(n) & \cdots & f_1(n)
\end{vmatrix} \tag{8}
$$

giving

$$B_0 = \tfrac{1}{2}, \quad B_1 = -\tfrac{1}{6}, \quad B_2 = B_4 = B_6 = \cdots = B_{2n} = 0,$$
$$B_3 = \tfrac{1}{30}, \quad B_5 = -\tfrac{1}{42}, \cdots.$$

We may now take another "triangle":

$$
\begin{array}{ccccccccccc}
 & & & & & 1 & & & & & \\
 & & & & 1 & & 1 & & & & \\
 & & & 1 & & 3 & & 1 & & & \\
 & & 1 & & 6 & & 7 & & 1 & & \\
 & 1 & & 10 & & 25 & & 15 & & 1 & \\
 & & & \cdot & \cdot & \cdot & \cdot & \cdot & & & \\
 & 1 & F_1(n-1) & F_2(n-1) & \cdots & \cdots & \cdots & F_{n-2}(n-1) & F_{n-1}(n-1) & & \\
1 & F_1(n) & F_2(n) & \cdots & \cdots & \cdots & \cdots & \cdots & F_{n-1}(n) & F_n(n) &
\end{array}
$$

where the n-th row is obtained from the preceding row by the expression

$$(n - i)F_i(n - 1) + F_{i+1}(n - 1) = F_{i+1}(n). \tag{9}$$

For example, from the third row: 1, 6, 7, 1, we obtain the fourth row: $1, 4\cdot 1 + 6 = 10$, $3\cdot 6 + 7 = 25$, $2\cdot 7 + 1 = 15$, 1. The following theorem is true for the $F_i(n)$:

THEOREM 3. *The elements in the $(n - 1)$st row are the coefficients in the expansion of x^n as a function of the factorials $x^{(i)}$.*

For example:

$$x^4 = x^{(4)} + 6x^{(3)} + 7x^{(2)} + x.$$

From the generating equation (9) we can obtain, as before, the form of the functions $F_0(n)$, $F_1(n)$, $\cdots$

$$\begin{aligned} \Delta F_{i+1}(n - 1) &\equiv F_{i+1}(n) - F_{i+1}(n - 1) = (n - i)F_i(n - 1) \\ F_{i+1}(n - 1) &= \Delta^{-1}[(n - i)F_i(n - 1)]. \end{aligned} \tag{10}$$

Since $F_0(n) = 1$ for all n

$$\begin{aligned} F_1(n - 1) &= \Delta^{-1} n = \frac{n^{(2)}}{2} \\ F_2(n - 1) &= \Delta^{-1}\left[(n - 1)\frac{n^{(2)}}{2}\right] = \frac{3n^{(4)} + 4n^{(3)}}{24} \\ F_3(n - 1) &= \Delta^{-1}\left[(n - 2)\frac{3n^{(4)} + 4n^{(3)}}{24}\right] \\ &= \frac{n^{(6)} + 4n^{(5)} + 2n^{(4)}}{48}. \end{aligned} \tag{11}$$

From these coefficients we can generate the numbers of Laplace (the numbers L_m below must be divided by $m!$ to yield the numbers of Laplace):

$$\begin{aligned} &L_1 = \tfrac{1}{2} \\ &L_1 + L_2 = \tfrac{1}{3} \\ &L_1 + 3L_2 + L_3 = \tfrac{1}{4} \\ &L_1 + 7L_2 + 6L_3 + L_4 = \tfrac{1}{5} \\ &L_1 + 15L_2 + 25L_3 + 10L_4 + L_5 = \tfrac{1}{6} \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ &L_1 + F_{n-1}(n)L_2 + F_{n-2}(n)L_3 + \cdots + L_{n-1} = \frac{1}{n + 1} \end{aligned} \tag{12}$$

giving

$$L_1 = \tfrac{1}{2}, \qquad L_2 = -\tfrac{1}{6}, \qquad L_3 = \tfrac{1}{4}, \qquad L_4 = -\tfrac{19}{30}, \qquad L_5 = \tfrac{9}{4}.$$

A determinantal solution is also obvious.

CALIFORNIA INSTITUTE OF TECHNOLOGY.

ON THE PROBABILITY OF ATTAINING A GIVEN STANDARD DEVIATION RATIO IN AN INFINITE SERIES OF TRIALS

By Joseph A. Greenwood and T. N. E. Greville

Suppose an event with constant probability p of occurrence to be repeated an infinite number of times, and suppose the ratio of the deviation from the expected number of successes to the standard deviation $\sqrt{npq}$ to be recomputed after each trial. We are interested in the probability that this ratio will at some time equal or exceed some positive number k. It is not difficult to show that the value of this probability is unity, but as the fact has not, to our knowledge, been previously pointed out in the literature, we give the following proof.

Let x_n denote the number of successes obtained in the first n trials, let

$$t_n = \frac{x_n - np}{\sqrt{npq}},$$

and let P denote the probability that, for some n, $t_n \geqq k$. We shall prove that $P = 1$. To do this, let the infinite series of trials be subdivided into consecutive, mutually exclusive subseries of finite length, and let m_i denote the number of trials in the i-th subseries. Let $N_i = \sum_{j=1}^{i-1} m_j$ for $i \geqq 2$, while $N_1 = 0$. Let k' be any number greater than k, and let m_i be so chosen that

$$m_i \geqq \frac{k'^2 p}{q} \qquad \text{for every } i, \tag{1}$$

and

$$\sqrt{m_i}\left(k' - k\sqrt{\frac{N_i}{m_i} + 1}\right) \geqq N_i \sqrt{\frac{p}{q}} \qquad \text{for } i \geqq 2. \tag{2}$$

It follows from (1) that

$$m_i \geqq m_i p + k' \sqrt{m_i pq} \qquad \text{for every } i. \tag{3}$$

It follows from (2) that

$$m_i p + k' \sqrt{m_i pq} \geqq (N_i + m_i)p + k\sqrt{(N_i + m_i)pq} \qquad \text{for every } i. \tag{4}$$

Let y_i denote the number of successes in the i-th subseries. It is evident from (4) that if

$$y_i \geqq m_i p + k' \sqrt{m_i pq} \tag{5}$$

for any i, then

$$t_{N_i + m_i} \geqq k.$$

Hence P is at least equal to the probability that (5) holds for some i.

Let p_i denote the probability that (5) holds for a particular i. It follows from (3) that, for every i, $p_i > 0$. Moreover, there exists a positive integer M

and a number $h > 0$, such that if $m_i \geqq M$, $p_i \geqq h$.[1] Since there is but a finite number of possible values of m_i less than M, there is a number $p_0 > 0$ such that $p_i \geqq p_0$ for every i. Hence the probability that (5) holds for no value of i is at most

$$\lim_{s \to \infty} (1 - p_0)^s = 0 .$$

Therefore, the probability that (5) holds for some i is unity.

DUKE UNIVERSITY
AND
UNIVERSITY OF MICHIGAN.

[1] Uspensky, J. V., *Introduction to Mathematical Probability*, p. 129.